Accepted Manuscript

Bid–ask spread estimator from high and low daily prices: Practical implementation for corporate bonds

Belén Nieto

PII: S0927-5398(18)30041-0
DOI: https://doi.org/10.1016/j.jempfin.2018.06.003
Reference: EMPFIN 1051

To appear in: *Journal of Empirical Finance*

Received date : 5 May 2017
Revised date : 8 June 2018
Accepted date : 11 June 2018

Please cite this article as: Nieto B., Bid–ask spread estimator from high and low daily prices: Practical implementation for corporate bonds. *Journal of Empirical Finance* (2018), https://doi.org/10.1016/j.jempfin.2018.06.003

*Separate Title Page

Bid–Ask Spread Estimator from High and Low Daily Prices: Practical Implementation for Corporate Bonds

Belén Nieto
belen.nieto@ua.es
Financial Economics and Accounting
University of Alicante
Address: Ctra. San Vicente sn, E-03690 Alicante, Spain
Phone: +34965903621

Highlights:

The Corwin and Schultz high–low volatility and spread measures are downward biased.

The bias is concentrated on assets that do not trade continuously daily.

The underestimation increases from the least to the most volatile assets.

This paper proposes a generalized version that accounts for days without any trade.

This proposal, once negative spread values are discarded, is more accurate.

Bid–Ask Spread Estimator from High and Low Daily Prices: Practical Implementation for Corporate Bonds

1. Introduction

Liquidity has become an important driver of asset prices. Specially, corporate bond returns have great exposure to liquidity risk. For example, during the 2008–2009 crisis, 30% of the variability in credit spreads could be explained by liquidity. The vast numbers of papers measuring liquidity and evaluating its effect on market prices are therefore not surprising.[1] Most of the liquidity proxies designed and previously estimated in the stock market have been directly translated to this market. However, the structure of the stock and bond markets is clearly different. Schestag et al. (2016; SSU hereafter) evaluate the appropriateness of a large set of daily liquidity proxies by comparing them with measures based on intraday data. These authors conclude that Corwin and Schultz's (2012; hereafter CS) transaction cost measure appropriately captures cross-sectional differences and time-varying patterns.

CS propose daily estimators for the two unobservable components of price changes, the bid–ask spread and volatility, which employ only high and low daily prices. The idea is to combine the information regarding a single day with a two-day period. Independently of the theoretical assumptions of their model, estimation of the CS measure presents practical concerns that are not addressed by SSU and which can be particularly relevant to corporate bonds. First, having a daily spread estimator requires that the asset be traded on all days and at least twice a day. This empirical requirement is hard to hold for corporate bonds, even if the most active ones are selected. Second, under the model, the variance in a two-day interval is twice the variance for one day in the interval. Again, this assumption is especially problematic in this market, since it is characterized by high volatility precisely in moments of low liquidity. Third, volatility and spread estimates depend on each other and can only be computed numerically with a high time cost, unless Jensen's inequality is ignored. This paper evaluates the effects that these practical issues can have on the accuracy of volatility and the spread estimates for corporate bonds.

With regard to the first point above, in the original paper of CS, the estimation is applied to the stock market, where the assumption of observable consecutive prices is reasonable. In their sample, on average, there is no trade in 4.11% of all working days. For these days without trades, the authors suggest the assumption that the high and low prices are the same as those observed the most recent prior trading day. This assumption imposes zero volatility, so that the final estimators understate the volatility and overstate the spread. In the sample of bonds used in this paper, the problem of infrequent trading affects 16% of trading days. To avoid the unrealistic imposition of zero volatility, I propose an adjusted version of the CS spread and volatility estimators that account for irregular intervals between two observable prices. Therefore, the adjusted estimators can be obtained without assumptions about prices on days without trades. This is the first contribution of this paper.

With respect to the second point above, as CS recognize, in periods of high volatility, the variance in a two-day interval can be more than twice the variance for a single day. In these cases, the resulting spread estimate is negative. The authors propose setting these negative daily values to zero since this approach produces more accurate estimates than other alternatives. In the sample of CS, negative spreads are obtained in 29.26% of the days, on average, across stocks. This percentage is similar in my sample of bonds. However, my sample includes the crisis period, which is characterized by highly volatile financial markets. Specifically, the volatility of daily bond returns is 1.86%, on average, between April 2007 and December 2009 and 1.04% during the remainder of the sample period. Therefore, spreads can be estimated as especially low or negative and the imposition of a zero spread against just omitting negative observations will reduce the average spread estimate. I compare the accuracy of the two approaches.

Finally, regarding the third point, I evaluate the differences in estimates between ignoring Jensen's inequality, which is computationally easy, and not ignoring it, which suffers from the high time cost of numerical computation.

With all of this in mind, this paper's second contribution is a detailed empirical analysis that allows for a clear practical implementation of the measure based on high and low daily prices in the case of corporate bonds (but also for other assets with non-continuous trading and non-constant volatility). I analyze the importance of the effect produced by the assumptions imposed in the standard estimation by evaluating the accuracy of the different estimators that relax these practical impositions. To do so, I compare the different CS-based volatility estimators with the realized volatility and the different CS-based spread estimators with three transaction cost benchmarks based on high-frequency data.

The first result is that the standard high–low volatility and spread estimators are significantly downward biased. On average, the CS volatility estimator is around 25% lower than the realized volatility in both time series and cross-sectional analyses. The CS spread estimator is 11% to 27% lower than the intraday proxy, depending on the benchmark, in the time series and even higher biases are obtained in the cross section. More importantly, both estimators show a problematic decreasing trend from the most active (liquid) bonds to less active bonds.

The volatility analysis shows that the negative bias is generally observed for all the bonds but monotonically increases in absolute value with the volatility level (percentage of days without trades). Jensen's inequality produces slightly larger estimates than the standard approach does and thus contributes to reducing the bias. However, the differences between the two are unremarkable. The use of the generalized version that accounts for non-continuous trading proposed here significantly reduces this bias and allows one to obtain volatility estimates that increase with effective volatility.

Regarding the spread, the standard CS proxy appears to be upward biased in the subsample that contains the 25% of the most liquid bonds but downward biased for the remaining bonds. The upward bias is eliminated when the expression corrects for the number of days between consecutive trades. The downward bias is considerably reduced when negative spreads are discarded instead of set equal to zero. The combination of the two adjustments reproduces reasonably well the time series and the cross-sectional distribution of the intraday proxies for the 50% of bonds in the central position of the sample in terms of trading frequency. As in the volatility estimation, if, in addition, Jensen's inequality is taken into account, the accuracy of estimates slightly improves. For the whole sample of bonds and, on average, for the three intraday proxies, the bias created by the standard methodology is -0.19 in the time series and -0.17 in the cross-section, in contrast to 0.05 and -0.03, respectively, obtained with the estimator that accounts for non-continuous trading and Jensen's inequality and discards negative spreads.

The remainder of the paper is organized as follows. Section 2 summarizes the CS proposal and its standard estimation in practice. Section 3 presents the adjusted estimator that accounts for infrequent trading and discusses alternatives that do not impose unrealistic assumptions. Section 4 describes the data. Section 5 compares the different estimation proposals with the standard one, while Section 6 evaluates their accuracy through their comparison with some benchmark proxies. Section 7 concludes the paper.

2. CS spread estimator

2.1. Volatility and bid–ask spread measures

CS propose an estimator based on the assumptions that the stock price follows a constant diffusion process and the daily high price (H) is a buyer-initiated trade and the daily low price (L) is a seller-initiated trade. Then, the log high–low ratio for observable (o) and true/actual (A) prices for day t are related as following:

$$ln\left(\frac{H_t^o}{L_t^o}\right) = ln\left[\frac{H_t^A(1+S/2)}{L_t^A(1-S/2)}\right] = ln\left(\frac{H_t^A}{L_t^A}\right) + \alpha, \qquad (1)$$

where $\alpha = ln[(2+S)/(2-S)]$ and S is the spread.

Additionally, CS assume that the spread is constant over two-day periods and the equation for the log of the high–low log ratio over the two days is then

$$ln\left(\frac{H_{t,t+1}^o}{L_{t,t+1}^o}\right) = ln\left(\frac{H_{t,t+1}^A}{L_{t,t+1}^A}\right) + \alpha. \qquad (2)$$

The square of equations (1) and (2) illustrates the main idea of the paper: the high–low price ratio has one component due to price volatility and another due to the bid–ask spread. The volatility is proportional to the data frequency, whereas the spread is not. Therefore, working simultaneously with two frequencies (one and two days), it is possible to identify (and estimate) the two components:

$$\alpha = -k_2\sigma + \sqrt{k_1\sigma^2(k_2^2 - k_1) + {}^{\beta}/_{2}} \qquad (3)$$

$$\sigma^2\left(k_2^2\left(2-2\sqrt{2}\right)+k_1\right)+\sigma k_2\left(2\sqrt{2}-2\right)\sqrt{\sigma^2\left(k_2^2-k_1\right)+{}^{\beta}\!/_{2}}+\frac{\beta}{2}-\gamma=0 \qquad (4)$$

where $k_1 = 4ln(2)$, $k_2 = \sqrt{8/\pi}$, σ is the volatility of each single day, β represents the expected value of the sum for two consecutive days of the square of the high–low log ratio, and γ is the expectation of the square of the high–low log ratio over the two days,

$$\beta = E\left\{\sum_{j=0}^{1}\left[ln\left(\frac{H_{t+j}^{o}}{L_{t+j}^{o}}\right)\right]^2\right\},\ \gamma = E\left\{\left[ln\left(\frac{H_{t,t+1}^{o}}{L_{t,t+1}^{o}}\right)\right]^2\right\}.$$

Equation (4) should be solved numerically and, introducing the solution in equation (3), the values for alpha and then for the spread are obtained.

2.2. Practical estimation

Both the assumptions underlying the theoretical measure and the computational cost of estimating the volatility numerically prompt CS to suggest some solutions for its practical implementation.

First, since the measure employs only daily data, one of its main advantages is ease of computation. However, this is offset by the high time cost that equation (4) requires to be solved numerically.[2] To eliminate this disadvantage, CS suggest ignoring Jensen's inequality. Thus $k_2\sigma = \sqrt{k_1}\sigma$ and closed-form solutions corresponding to (3) and (4) are possible:

$$\alpha = \frac{\sqrt{2\beta}-\sqrt{\beta}}{3-2\sqrt{2}} - \sqrt{\frac{\gamma}{3-2\sqrt{2}}}, \qquad (5)$$

$$\sigma = \frac{\sqrt{\beta/2}-\sqrt{\beta}}{\left(3-2\sqrt{2}\right)\sqrt{k_1}} + \sqrt{\frac{\gamma}{\left(3-2\sqrt{2}\right)k_1}}, \qquad (6)$$

which are estimated each day t using

$$\hat{\beta} = \left[ln\left(\frac{H_{t-1}}{L_{t-1}}\right)\right]^2 + \left[ln\left(\frac{H_t}{L_t}\right)\right]^2 \text{ and } \hat{\gamma} = \left[ln\left(\frac{H_{t-1,t}}{L_{t-1,t}}\right)\right]^2.$$

Second, the estimator assumes that the asset trades continuously on all days and at least twice a day. Therefore, in practice, infrequent trading is a problem that must be addressed. In their original paper, CS adopt the following solution. For those days when the asset shows only one trade, they propose an adjustment that allows estimating different high and low prices and, for cases of no trades during a day, the high and low prices are assumed to be the same as those observed the most recent prior trading day. This assumption imposes $\hat{\gamma} = \hat{\beta}/2$ for days without trades, which produces a zero value for volatility and, consequently, overestimates the spread.

Third, the assumption of a constant diffusion process for the asset price implies that the volatility is proportional to the interval and thus the true variance over a two-day interval is twice the variance over a single day. Although this assumption could be acceptable, on average, clearly it does not hold day to day. In practice, the assumption can be broken when the midpoint for one day is far from the midpoint for the other day or when the distance between the high and low prices is different enough when comparing the two days. In these cases of high volatility, one can find $\hat{\gamma} > \hat{\beta}$ (contrary to the theory) and thus $\hat{\alpha} < 0$.

3. Generalized estimator and practical issues in the estimation procedure

The aim of this paper is twofold. On the one hand, I evaluate the effect of the assumptions described above, which are typically imposed in the practical estimation of CS's volatility and transaction cost measures, for the corporate bond market. On the other hand, I propose alternative practices that can improve the accuracy of the estimates, one of which generalizes the measures to avoid the assumption of continuous trading.

3.1. Infrequent trading problem: A generalized estimator that accounts for non-consecutive trading days

Corporate bonds and other assets are not traded continuously daily and thus high and low prices are not observable for all days in a sample period. As stated, the imposition of high and low prices equal to those of the prior day underestimates volatility and overestimates the spread. My proposal avoids this imposition. I generalize the measure to also be applied when trades are non-consecutive, such that volatility and the bid–ask spread are estimated only for days with trades and the estimates incorporate the effect of a time gap between two observable prices.

The foundation of my proposal is the same as for CS: I combine information at a daily frequency with information at a lower frequency; however, this other frequency changes each day and is defined by the observation of prices. The following steps in the derivation of the estimates are along the lines of the original CS paper.

Following equation (1), the square of the high–low ratio for day t is

$$\left[ln\left(\frac{H_t^o}{L_t^o}\right)\right]^2 = \left[ln\left(\frac{H_t^A}{L_t^A}\right)\right]^2 + 2\alpha ln\left(\frac{H_t^A}{L_t^A}\right) + \alpha^2. \qquad (7)$$

Assuming that it is possible to have trades that are not on consecutive days, with n - 1 days between the current trade and the previous one, I define the sum of (7) over n single days in the period as follows:

$$\sum_{j=0}^{n-1}\left[ln\left(\frac{H_{t+j}^o}{L_{t+j}^o}\right)\right]^2 = \sum_{j=0}^{n-1}\left[ln\left(\frac{H_{t+j}^A}{L_{t+j}^A}\right)\right]^2 + 2\alpha\sum_{j=0}^{n-1} ln\left(\frac{H_{t+j}^A}{L_{t+j}^A}\right) + n\alpha^2. \qquad (8)$$

Taking the expectations in (8) yields

$$E\left\{\sum_{j=0}^{n-1}\left[ln\left(\frac{H_{t+j}^o}{L_{t+j}^o}\right)\right]^2\right\} = \sum_{j=0}^{n-1} E\left\{\left[ln\left(\frac{H_{t+j}^A}{L_{t+j}^A}\right)\right]^2\right\} + 2\alpha\sum_{j=0}^{n-1} E\left[ln\left(\frac{H_{t+j}^A}{L_{t+j}^A}\right)\right] + n\alpha^2. \qquad (9)$$

Garman and Klass (1980) and Parkinson (1980) show that, under the constant diffusion assumption, an estimator for the variance of the price change in a time interval can be obtained using only the distance between the maximum and minimum prices observed in the interval. The advantage of this estimator is that one does not need to measure the continuous price sample path during the interval; however, it is more efficient than the standard estimator that uses closing prices, for example. Using the derivation of the probability distribution for the distance between the maximum and minimum prices of Feller (1951), Parkinson (1980) shows that the moments of the distance are proportional to the variance. In particular, the two first moments are

$$E\left\{\left[ln\left(\frac{H_t^A}{L_t^A}\right)\right]^2\right\} = k_1\sigma^2 \text{ and } E\left[ln\left(\frac{H_t^A}{L_t^A}\right)\right] = k_2\sigma. \qquad (10)$$

Introducing equations (10) into (9) yields

$$E\left\{\sum_{j=0}^{n-1}\left[ln\left(\frac{H_{t+j}^O}{L_{t+j}^O}\right)\right]^2\right\} = nk_1\sigma^2 + 2\alpha nk_2\sigma + n\alpha^2. \qquad (11)$$

The left-hand side of (11) is denoted $\beta_{(n)}$ and thus the expression is

$$nk_1\sigma^2 + 2\alpha nk_2\sigma + n\alpha^2 - \beta_{(n)} = 0. \qquad (12)$$

From equation (12) we can solve for α. Since it is a positive transformation of the spread, α must be positive and thus I take the positive root:

$$\alpha = -k_2\sigma + \sqrt{\sigma^2(k_2^2 - k_1) + {}^{\beta_{(n)}}/_{n}} \qquad (13)$$

On the other hand, the square of the high–low ratio from the n-day period is

$$\left[ln\left(\frac{H_{t,t+n-1}^O}{L_{t,t+n-1}^O}\right)\right]^2 = \left[ln\left(\frac{H_{t,t+n-1}^A}{L_{t,t+n-1}^A}\right)\right]^2 + 2\alpha ln\left(\frac{H_{t,t+n-1}^A}{L_{t,t+n-1}^A}\right) + \alpha^2. \qquad (14)$$

Again, taking the expectations in (14), incorporating the volatility proportionalities given in (10), and denoting the left-hand side as $\gamma_{(n)}$ yields

$$nk_1\sigma^2 + 2\alpha\sqrt{n}k_2\sigma + \alpha^2 - \gamma_{(n)} = 0. \qquad (15)$$

Introducing the solution (13) for α into (15) and rearranging terms, we have

$$\sigma^2\left(k_2^2\left(2-2\sqrt{n}\right) + (n-1)k_1\right) + \sigma 2k_2\left(\sqrt{n}-1\right)\sqrt{\sigma^2(k_2^2-k_1) + {}^{\beta_{(n)}}\!/_{n}} + \frac{\beta_{(n)}}{n} - \gamma_{(n)} = 0$$

(16)

Using the expression (16), it is possible to solve σ numerically. Then, the solution is introduced into (13) to obtain the estimation for α.

Moreover, if Jensen's inequality is ignored, we can obtain closed-form solutions corresponding to expressions (13) and (16)

$$\alpha^* = \frac{\sqrt{n\beta_{(n)}} - \sqrt{\beta_{(n)}}}{n+1-2\sqrt{n}} - \sqrt{\frac{\gamma_{(n)}}{n+1-2\sqrt{n}}}, \qquad (17)$$

$$\sigma^* = \frac{\sqrt{\beta_{(n)}/n} - \sqrt{\beta_{(n)}}}{\left(n+1-2\sqrt{n}\right)\sqrt{k_1}} + \sqrt{\frac{\gamma_{(n)}}{\left(n+1-2\sqrt{n}\right)k_1}}. \qquad (18)$$

Equations (17) and (18) match the standard CS estimators (equations (5) and (6)) if the asset is traded daily continuously and n = 2 for all the days during the asset's trading life. However, my estimators can also be used for the general case of n - 1 days between trades. In a practical application, the n-day period is defined by the availability of prices. Therefore, on the one hand, my proposal does not impose estimation of the spread and volatility for all the days. If there is no trade one day, the estimators are not computed. As in the standard case, the volatility estimator will be zero for those days in which $\hat{\gamma}_{(n)} = \hat{\beta}_{(n)}/2$. However, this will occur when both high and low observable prices for the two extreme days in the n-day period are effectively equal. In contrast, the standard volatility estimator would be imposed to be zero for days without trades. This practice will understate the volatility estimator for assets or periods of low trading activity. Consequently and inversely, the spread proxy would be upward biased.

Additionally, for days with observable prices, if $n > 2$, both the spread and the volatility-adjusted estimators are lower than the standard one because they take into account the distance between trades. The two effects produce a positive difference between the standard CS spread estimator and the adjusted one.

It must be clear that the adjusted estimators are subject to the same criticisms as the standard ones, in the sense that the model relies on the same theoretical assumptions. The adjustment here constitutes an improvement for their practical estimation, because it avoids unrealistic assumptions about non-observable prices.

3.2. Negative spreads

CS find that the mean values of the simulated spread estimates without adjustment for negative cases are very closed to the true spreads when the simulation is done under ideal conditions. However, they are clearly lower than the true spreads when the simulation incorporates infrequently observed prices and overnight returns. They compare the performance of two alternative estimators that either set zero or delete negative values, computed with daily stock data from CRSP and the period 1993 to 2006, with the effective spread computed using NYSE TAQ data. They find that the mean and the median are closer for the first alternative. Obviously, the second alternative produces larger spreads.

A negative value for alpha can be obtained during days or periods of high volatility in which the theoretical assumption $\beta \geq \gamma$ does not hold. Therefore, generally, imposing zero spread instead of deleting negative daily estimates, reduces the average spread and particularly for assets and/or periods with high volatility prices. Moreover, the probability that this does occur will be higher for less frequently traded assets, since these assets usually show a higher level of volatility in prices. As well known, there is a positive correlation between volatility and illiquidity. Thus, this practice will produce a reduction in the estimated spread precisely for those assets and periods for which the spread should be larger. In fact, in their Internet Appendix, CS show that the estimator is downward biased for the full sample of stocks and this underestimation is concentrated in the smallest stocks, which have higher volatility and a lower trading frequency.[3]

I will evaluate how problematic it can be to impose a zero spread in cases of negative estimates for corporate bonds by comparing these results with those obtained from just omitting negative spread observations. Of course, in this second case, the number of days with estimates is lower but the mean spread for the asset will not be artificially reduced.

The problem of negative values also applies to the adjusted estimator; alpha in equation (17) is negative if $\gamma_{(n)} > \beta_{(n)}$. Therefore, I will first analyze separately the problems of infrequent trading and negative spreads. Second, I will investigate the combined effect of these two opposing potential biases.

3.3. Jensen's inequality

Finally, I also evaluate whether ignoring Jensen's inequality can suppose significant bias in both the volatility and the spread estimators for the sample of corporate bonds. The conditions that produce either unreal zero volatility or an unreal zero spread also apply to solutions that do not ignore the inequality. Therefore, I will analyze the differences between ignoring the inequality and not in the three cases: the standard measure, the adjusted measure that accounts for non-continuous trading, and when discarding the negative values for the spread.

4. Data

The initial sample consists of intraday U.S. corporate bond transaction data from the Trade Reporting and Compliance Engine (TRACE) and the corresponding bond characteristics from the Fixed Income Securities Database (FISD), all provided by Mergent, from July 2002 to December 2014. I apply the filters described by Dick-Nielsen (2009, 2014) to clean duplicates, corrections, and reversals and a median filter with five standard deviations computed with a previous window of 90 natural days is used to eliminate extreme outliers and erroneous reports.[4]

From the intraday frequency data, I construct a daily sample. For each bond and day, I record the high, low, and last prices; the number of trades within the day; and the cumulative trading volume at the end of the day. Requirements are then applied. I include only bonds with FISD information. I delete trades on holidays or days outside of the bond's life, defined by the offering and maturity dates available from FISD. Additionally, bonds in default are only included up to three months before the default date and after the reinstated date, if any. This approach provides a final sample of 89,654 bonds. I complete the daily information with the rating associated with each bond and day using the historical rating changes for each of the four rating types available from FISD: Standard & Poor's, Moody's, Fitch, and DP Information Group. Panel A of Table 1 provides the mean, standard deviation, and three quartiles (Q1, Q2, and Q3) of some characteristics of this complete and clean database.

To make comparisons, I apply the same sample selection criteria as SSU do: The period is from October 1, 2004, to September 30, 2012, and the bonds must show at least one year of active trading and must be traded on at least 75% of the trading days. This selection produces a sample of 3,526 bonds.[5] Panel B of Table 1 contains the descriptive statistics of these selected bonds. Comparing panel B of Table 1 with the results in Table 1 of SSU shows the similitude of the samples. My selected bonds have lower dispersion in maturities, are slightly better rated, and show a slightly higher number of trades per bond.[6] Comparing panels A and B and looking at the median (Q2), for example, reveals selected bonds to be more active, as expected. They show an average of 7.58 trades per day (2.56 in the entire sample) and a total of 6,677 trades (61 in the entire sample) and are traded on 91.31% of the days (17.52% in the entire sample). Additionally, the offering amount is considerably larger for these more active bonds. Other differential characteristics in the selected sample regarding the full universe are the following: bonds have longer maturities, larger coupons, worse rating classifications, and a slightly higher Treasury spread and more than the 50% of the bonds are issued by industrial firms.

To analyze and compare the different estimate proposals, the sample of bonds is split into four subsamples regarding the number of days without trades (observable prices). Specifically, for each bond, I compute the average number of days between trades: the minimum is one for bonds that are traded all days during their lifetime and the maximum is 1.3361. I use the quartiles of the distribution of this variable to split the sample so that the four subsamples contain the same number of bonds (882 in subsamples 1 and 3 and 881 in subsamples 2 and 4) but the average number of days between trades is below 1.0252 for subsample 1, ranges from 1.0252 to 1.0948 for subsample 2, ranges from 1.0948 to 1.2015 for subsample 3, and is above 1.2015 for subsample 4. This split is appropriate to analyze not only potential bias related to infrequent trading but also bias due to sufficiently high volatility. As shown in Table 2, there is a monotonic inverse relation between volatility and trading frequency.

The four panels in Table 2 present descriptive statistics for the subsamples. For example, the bonds in subsample 1 have a very low value for the number of days between trades, 1.01, on average, with a very low standard deviation as well. Looking at the median values, one sees they are traded 17.57 times per day, on average, with a trading volume of 5,577,183 USD per day and a total of 15,618 trades during their lifetime. In a comparison among panels, the activity indicators decrease from panel A to panel D. The number of days with a trade, the average number of trades per day, the total number of trades in the bond's life, and the trading volume decrease monotonically from subsample 1 to subsample 4. Additionally, relations between bond characteristics and liquidity patterns are observed; more active bonds have lower maturity, a much higher offering amount, lower coupons, and better rating classifications and are issued in higher proportions by financial institutions. Finally, I estimate the realized volatility for each bond using its complete lifetime and daily returns computed from closing prices. The last row of Table 2 presents descriptive statistics for the average volatility within each subsample. As expected, a lower trading frequency is associated with higher volatility, reflecting the negative relation between liquidity and volatility.

5. Estimation results

This section compares the results between the different estimators and practical assumptions discussed previously. The standard and the non-continuous trading adjusted versions, accepting the Jensen inequality or not, are compared for both components of the high-low ratio, the spread and volatility. Additionally, in the case of the spread, I evaluate the effect of setting the negative values to zero against treating them as missing values. Results for volatility estimates are in Section 5.1 and panel A of Table 3, and those for the spread estimates are in Section 5.2 and panel B of Table 3.

5.1. Volatility estimates

In relation to the infrequent trading issue, the standard practical application of the CS methodology imposes zero volatility in those days without trades. Therefore, the standard volatility estimator is expected to be downward biased for assets that are non-continuously traded and in periods of low activity. To evaluate the effect of such an imposition, I compare the standard estimator (Standard) in equation (6) with the adjusted proposal (Adjusted) in equation (18), which does not produce any estimates on days without trades and incorporates the effect of more than one day between trades. Additionally, I evaluate whether ignoring Jensen's inequality has relevant consequences in the final volatility estimators in both cases, the standard and the adjusted, by using equations (4) and (16) (Jensen and Adjusted Jensen, respectively).[7] These four estimates are computed daily and for each bond. Finally, I count the percentage of days for which the estimated volatility is zero.

The analysis is based on the cross-sectional differences that can be found when comparing the different estimates. Thus, I work with the mean value of daily volatility for each bond and estimator and compute the relative differences (Diff.) between the standard and each of the other alternative estimators. Panel A of Table 3 reports the average results across the bonds in each subsample. All values are in percentages. The first row shows that the standard volatility estimate is around 1% and, contrary to expectations, it displays a decreasing pattern from subsample 1 to subsample 4. This pattern corresponds to an increasing trend in the percentage of days in which the resulting estimate equals zero. This percentage is only 2% when the average number of days between trades is lower than 1.0252 (subsample 1), while a large value of 29% can be achieved for bonds with at least 1.2015 days between trades, on average (subsample

4). For the whole sample of bonds, the problem of infrequent trading affects 16% of the days, in contrast to 4.11% in the sample of stocks used by CS. The second two rows display results regarding the adjusted estimator. As expected, without the imposition of zero volatility on days when prices are not observable, the percentage of days with zero volatility is considerably lower, especially for bonds that are less traded. Now cases of zero volatility correspond to days in which the high and low prices effectively coincide with those of the previous trading day. Consequently, the adjusted estimator is higher than the standard one for the four groups of bonds. Moreover, the estimator shows a quasi-increasing pattern from the most active bonds to the less active group. The relative difference between the standard and adjusted estimates is negative, going from -0.93% for subsample 1 to -19.47% for the less traded bonds, and is statistically significant for subsamples 3 and 4.[8] Panel A of Table 3 also shows that the incorporation of Jensen's inequality has a positive effect on volatility estimates but the effect decreases as the frequency of trading decreases. In any case, differences between Standard and Jensen estimates are not statistically relevant. In the same sense, results for the two adjusted versions, with and without the Jensen's inequality, are similar each other.

The time series of the standard and adjusted estimators in the two extreme subsamples are displayed at the top and bottom of Figure 1. I compute monthly volatility estimates for each bond by averaging daily estimates among the days within the corresponding month. Then, monthly aggregate estimators for each subsample are obtained as the average across all bonds within the subsample. Figure 1 illustrates the downward bias of the standard CS volatility estimator and the increase in distance between the standard and adjusted approaches when the frequency of trading decreases.

5.2. Spread estimates

Panel B of Table 3 reports the results regarding the mean value of the spread estimates. As in Panel A, the standard proxy, the proxy that incorporates Jensen's inequality, and the two corresponding adjusted measures are compared. In addition, in this case, I also evaluate the effect of two possible adjustments for negative daily spread estimates (days with high enough volatility): either negative values are set to zero or excluded. In the second case, estimators are denoted by "Non-zero". The term *% zeros* refers to the percentage of days in which the estimated spread is negative.[9] The different combinations give a total number of eight possible spread estimators. In all cases, the relative differences are computed as the standard proxy minus the alternative one. Figures 2 and 3 represent the time series of some monthly aggregate spread estimators obtained by averaging across all bonds in subsamples 1 and 4, respectively.

The first row in panel B of Table 3 shows that the standard CS estimator produces a global mean (for the whole sample of bonds) of 0.77% and the spread is more or less decreasing with the bonds' trading frequency, going from 0.89% in subsample 1 to 0.72% in subsample 4.[10] The estimation of the adjusted version, without the imposition of zero volatility for days without trade, produces lower values confirming the upward difference of the standard measure. The mean value across all bonds in subsample 1 is similar between the two estimators, with a small and non-significant relative difference of 2.5%. However, the differences increase monotonically with the proportion of days without trades, are significantly different from zero for subsamples 2 to 4, and can achieve a large value of 30.7% for the less traded bonds. Panel A of Figure 2 shows that the difference between the standard and adjusted estimators is practically nonappreciable when the bond trades quasi-continuously, whereas it is noticeable in the case of bonds with an average gap of 1.289 days between trades (panel A of Figure 3). Jensen's inequality makes the spread estimates slightly higher than the standard ones. As in the volatility analysis, the highest effect is for bonds in subsample 1, although relative differences between Standard and Jensen are not significant, in any case.

All the three estimators (Standard, Adjusted and Jensen) produce the counterintuitive result that the spread is larger for the most active bonds than for the others. This result can be driven by two facts. On the one hand, the theoretical model behind the estimators implies that the spread is low when volatility is high, for a fixed high–low ratio. Given the negative association between volatility and liquidity (frequency of trading) displayed in Table 2, the decreasing pattern in spreads from subsample 1 to subsample 4 could be expected. On the other hand, setting to zero instead of discarding negative values contributes to magnifying the negative correlation between spread and volatility. Again, the lower the liquidity, the higher the volatility and the greater the number of days for which the spread cannot be reasonably estimated. As indicated in the second row of panel B of Table 3, this occurs in 19% of the days in the group of the most active bonds and in 32% in subsamples 3 and 4.[11] If we compute the time series correlation between daily standard spread and volatility estimators, we find that it is negative for all bonds and increases in absolute value with bond volatility, from -22.5% in subsample 1 to -28.8% in subsample 4. Similar correlations are obtained with the estimates that incorporate Jensen's inequality. The infrequent trading adjustment reduces the level of this correlation but still produces values of -21.7% or -23.8% for subsamples 1 and 4, respectively. Therefore, discarding negative values will increase the spread estimate, on average, on the one hand, and reduce the undesirable decreasing pattern in the spread as volatility (and illiquidity) increases, on the other hand.

Rows seven, nine and eleven in Panel B of Table 3 display the results for the Standard, Adjusted and Jensen average spread estimators discarding negative daily values. Comparing the Standard with Standard Non-zero, the second spread is larger for all the bonds, with a much higher global mean value of 1.04%, and the decreasing pattern disappears. The differences are significantly large for the four subsamples, and especially for the less liquid bonds. The graphs in panel B of Figures 2 and 3 illustrate this conclusion.

The combination of the infrequent trading adjustment with the elimination of days with negative spreads (Adjusted Non-zero) makes the two opposite effects compensate for each other, but differences regarding the standard measure are still observable. After the two adjustments, one finds a final negative and significant difference for the four subsamples. As Table 3 and panels C of Figures 2 and 3 indicate, the magnitude of the difference between Standard and Adjusted Non-zero estimators is similar between subsamples. The two adjustments cause spreads to be greater than the standard one during normal times but the differences are close to zero during the crisis period. Finally, if, in addition, Jensen's inequality is considered, the spreads are similar but a bit wider than in the previous case (panel D of Figures 2 and 3).

6. Comparison with volatility and spread benchmark measures

As Table 3 reveals, the different versions of the CS volatility and spread proxies can display very different values. To determine which combination of the modifications proposed here act appropriately, in this section, I evaluate the accuracy of the different CS-based estimators.

The four CS volatility estimators (Standard, Adjusted, Jensen and Adjusted Jensen) are compared with the realized volatility estimated each day and for each bond using daily returns computed form closing prices in a previous window of three months.[12] To ensure a fair comparison, daily CS volatility estimates are then replaced by their rolling average in the last three months.

In the case of the spread, the eight CS-based estimators (Standard, Adjusted, Jensen, Non-zero, Adjusted Non-zero, Jensen Non-zero, Adjusted Jensen, and Adjusted Jensen Non-zero) are compared with three alternative intraday transaction cost proxies. The three intraday measures are selected on the basis of the availability of the data set. Two of the measures infer the bid–ask difference from the distribution of trade prices.

The first one is the well-known proposal of Roll (1984) applied, in this case, to intraday returns:

$$Roll_t = \begin{cases} 2\sqrt{-Cov(r_{t,s}, r_{t,s-1})} & if\ Cov(r_{t,s}, r_{t,s-1}) < 0, \\ 0 & otherwise, \end{cases}$$

where $r_{t,s}$ is the intraday return computed with two consecutive trades represented by the times *s* - 1 and *s* within each day *t*. The covariance is computed daily only for days with at least four trades and using intraday returns starting from the second trade. The higher the negative autocovariance, the larger the spread.

The second measure is the interquartile range of trade prices within a day of Han and Zhou (2007) and Pu (2009):

$$IQR_t = \frac{P_t^{Q0.75} - P_t^{Q0.25}}{\bar{P}_t},$$

where $\bar{P}_t$ is the average price on day *t*. This measure is computed only for days with at least three trades. The larger the interquartile range, the greater the intraday volatility, which, for corporate bonds, is an indirect indicator of large spreads.

The last proxy is the round-trip cost measured by Feldhütter (2012). Bonds are commonly observed to trade multiple times with the same volume during a short time interval. These cases can be interpreted as round-trip trades that dealers undertake to put buys and sells together and then earn on the spread. Therefore, the spread can be estimated by working only with the prices involved in round-trip transactions. Following SSU, I use 15-minute intervals to identify round-trip trades and, for each set of round-trip trades, the cost is estimated as

$$Round_s = \frac{2(H_s - L_s)}{(H_s + L_s)/2},\ if\ H_s \neq L_s\ ,$$

where s is the number of round-trip trades identified in a day and H_s is the highest price and L_s is the lowest price within each set of round-trip trades. The daily measure is computed as the average of all round-trip cost estimates within the day.

The comparison is carried out in terms of the moments of the distribution, the correlation, and the bias between pairs of measures and both in time series and cross-sectionally, for each subsample of bonds based on the frequency of trading. Results in panel A of Tables from 4 to 9 refer to volatility analysis while the evaluation of the spreads is presented in panel B.

6.1. Time series analysis: Aggregate measures across subsamples

This case involves a time series comparison and I work with aggregate measures obtained by averaging individual measures across all bonds in each subsample.

Panel A of Table 4 provides descriptive statistics (mean, standard deviation and the three quartiles) for the time series distribution of the realized volatility, in the first row, and the four CS volatility estimators. As expected, realized volatility is negatively related to the frequency of trading (from 1.20% to 1.60%, on average, and from 0.94% to 1.36%, on median). For the most traded bonds, all CS estimates show a downward biased in mean and the third quartile (Q3) while they can well approximate the median and the left hand side of the distribution. For this subsample, the four estimators show similar values but the distribution of the adjusted version with the Jensen's inequality is closest to that of the realized volatility. For the other subsamples, all four estimators produce significantly smaller values than realized volatilities but differences are considerably lower for the two adjusted versions. Therefore, the adjustment for non-continuous trading contributes to improve the estimation of volatility. This conclusion can be visualized in Figure 4 that displays aggregate (market-wide) series for the whole sample of bonds of the standard and adjusted volatility estimators and the realized volatility.

The first three rows of panel B of Table 4 provide descriptive statistics for the distribution of the intraday transaction cost proxies for each subsample and Figure 5 illustrates the time series of the corresponding aggregate measures for the whole sample of bonds. The global time series correlation between them is very high and goes from 0.93 between *IQR* and *Round* to 0.98 between *Roll* and *IQR*. The measures *Roll* and *IQR* produce similar mean and quartiles for subsamples 2 to 4 while *IQR* is lower than *Roll* for the subsample with the most active bonds. The measure *IQR* is more volatile than *Roll* in the whole sample and, as seen in Figure 5, it captures the increase in spreads during the crisis period with higher intensity. The measure *Round* produces much higher values for the complete distribution than the other two proxies for the four subsamples. However, as shown in Figure 5, *Round* is greater than the other proxies in normal times but lower during recessions. More importantly, all the three proxies display the expected increasing pattern as the frequency of trading decreases.

The remaining rows in panel B of Table 4 refer to the eight CS-based spread estimators. For all four subsamples, we see a clear difference between estimators that discard negative values and those that impose a zero spread in such cases. Generally, the four non-zero estimators show higher values than the intraday proxies for the subsample of bonds that trade at a minimum frequency of 1.0252 days, on average, while excluding negative spreads seems to be necessary to reasonably estimate transaction costs when the frequency of trading is lower. The standard CS estimator is upward biased for bonds in subsample 1 regarding *Roll* or *IQR* but, for all statistics, produces lower values than any of the three measures based on intraday information for subsamples 2 to 4. This result suggests a downward bias for the whole sample, confirming the result of SSU. In relation to *IQR* and *Roll*, the estimators that produce the most similar moments are Adjusted for subsample 1 and the two estimators that use the non-continuous trading adjustment and treat negative spreads as missing values for subsample 2. For the remaining bonds, the four Non-zero estimators show the moments of the distribution to be close to those observed for the intraday proxies, although they are upward biased in relation to *IQR* when the adjustment for infrequent trading is not considered.

Table 5 displays the time series correlations between the aggregate estimators for each subsample. The correlations between the four CS-based volatility estimators and the realized volatility are in panel A. They all are high due to the overlapping procedure and very similar to each other. In any case, the correlations are higher for the estimators that incorporate the infrequent trading adjustment for subsamples 2 to 4. Regarding spreads, in panel B, we observe first that the correlation between any CS-based estimator and any intraday proxy decreases with bond trading frequency. Second, the time pattern of any CS-based estimator is closer to that of *IQR* than to that of *Roll* or *Round*, for all subsamples. All correlations are higher than 75% when the benchmark is *IQR*. Third, for subsample 1, Adjusted Non-zero and Adjusted Jensen Non-zero show the highest correlation with both *Roll* and *IQR* and Adjusted and Adjusted Jensen both show the highest correlation with *Round*. For the group of bonds in subsample 2, the highest correlations are observed between Non-zero or Jensen Non-zero and *Roll* or *IQR* and between Adjusted or Adjusted Jensen and *Round*. For the less traded bonds in subsamples 3 and 4, the estimators that discard negative spread values generally show the highest correlations.

Lastly, I compare the mean and the three quartiles of the time series distribution of the different volatility and spread estimators in terms of the bias, defined as the difference between each of the CS estimators and the benchmark, in percentages. Results are displayed in Table 6. Panel A refers to volatility estimators and panel B refers to spread estimators and is organized in three blocks in relation to each intraday transaction cost proxy. Asterisks indicate that the null of equal means or equal quartiles is rejected at the 1% significance level. For comparison of the means, I use the Wilcoxon rank sum test with the null that the two samples come from identical continuous distributions with the same mathematical expectation. The comparison between the three quartiles is carried out with Pearson's chi-squared test. The null is that the frequency distribution in the observed samples is consistent with a theoretical distribution. In the case of the two samples, the statistic is

$$\sum_{i=1}^{2}\frac{\left(O_i^{<q}-E_i^{<q}\right)^2}{E_i^{<q}}+\sum_{i=1}^{2}\frac{\left(O_i^{\geq q}-E_i^{\geq q}\right)^2}{E_i^{\geq q}},$$

where O_i is the observed frequency for sample i and E_i is the expected theoretical frequency for values lower than q and higher or equal than q. The expected frequency is estimated with the values of the two samples simultaneously and q indicates the quantile, that is, 0.25, 0.5, or 0.75. Under the null, the difference between the observed and expected frequencies for the two samples is zero and the statistic has a chi-squared distribution with one degree of freedom.

As expected, given the previous results in panel A of Tables 3 and 4, Table 6 shows that CS volatility measures underestimate the realized volatility unless the bonds trade very frequently. In subsample 1, the four estimators are statistically unbiased for the median and Adjusted Jensen produces the lowest bias for all the distributional statistics. For the other subsamples, the CS-based estimators are negatively biased but considerably better results are obtained if the estimator accounts for the non-continuous trading problem.

In relation to spreads, taking *Roll* as the benchmark, for subsample 1, we see that measures that discard negative values overestimate the spread while the other measures produce accurate estimates. The lowest biases for the mean and the median are observed for Adjusted. In contrast, for subsamples 2 to 4, imposing a zero spread in case of a negative value generally produces a downward bias. The unique estimators that show biases indistinguishable from zero are Adjusted for the mean, Jensen Non-zero for the mean and the median, and Adjusted Jensen Non-zero for the median and Q3 in subsample 2 and Adjusted Jensen Non-zero for the mean in subsample 3. Finally, although statistically significant, Jensen Non-zero produces the lowest biases for the three quartiles of the distribution for bonds in subsample 4.

Regarding *IQR*, I find that all CS-based estimators produce positive and significant biases for frequently traded bonds. As in the *Roll* case, in this first subsample, the lowest biases are found for the estimator that adjusts for the period between trades. For subsample 2, imposing a zero spread when it is negative always creates a downward bias, while discarding negative values creates an upward bias. In any case, considerably lower biases are found for Adjusted Non-zero. For subsample 3, the most accurate estimator is the one that accounts for days without trades when negative cases are deleted and

Jensen's inequality is taken into account, which shows a zero bias for the mean and the two first quartiles.

In relation to *Round*, as expected, given the descriptive statistics shown in Table 4, all eight CS-based estimators are significantly downward biased for the bonds in subsamples 2 to 4 and the lowest biases are for Jensen Non-zero. Additionally, for subsample 1, Jensen and Adjusted Jensen produce the lowest biases for the mean and Q3, respectively.

Finally, it can be observed that the estimator adjusted by the number of days between trades and discarding negative values creates biases lower than in standard practice for subsamples 2 to 4 and for all three intraday benchmarks. For example, in subsample 2, with the biases for the mean, Adjusted Non-zero reduces the absolute bias in relation to Standard by 93%, 62%, and 58% when *Roll*, *IQR,* or *Round* is used as the benchmark, respectively. Equivalently, the reductions are 66%, 79%, and 43% in subsample 3 and 47%, 56%, and 30% in subsample 4, respectively. Moreover, if, in addition, Jensen's inequality is considered, these reductions are even greater.

6.2. Cross-sectional analysis

To analyze the cross-sectional distributional properties of the volatility and spread estimators, I now work with a monthly frequency by averaging each measure along all daily values within a month. Then, for each month in the sample period, I compute the cross-sectional mean, standard deviation, and three quartiles. Table 7 displays the average across all bonds in each subsample of these cross-sectional descriptive statistics. Comparing the results with those of Table 4, one sees that the mean and median values are lower than in the time series, but there is much greater cross-sectional dispersion for all the measures.

Conclusions about the similitude between the distribution of CS-based estimators and the benchmark measures are the same as in the time series analysis. The distribution of CS volatility is closest to that of realized volatility when the estimator accounts for gaps in trading, which allows estimates that increase as the frequency of trading decreases. Regarding spreads, adjusted versions show a distribution similar to that of *IQR* and *Roll* for subsample 1, although all the measures overestimate the third quartile in this subsample of highly active bonds. For the remainder of the sample, versions that discard days with a negative spread estimation better reproduce the distribution of the intraday measures. The descriptive statistics are similar to those of *Roll* and *IQR* and lower than those of *Round*.

Table 8 displays average values of the cross-sectional correlation between each pair of measures computed each month. As expected, cross-sectional correlations are lower than time series correlations for both volatilities and spreads. In panel A, we see, as before, that the correlation between the CS estimators and realized volatility is higher for the two adjusted versions and for all bonds. Curiously, the correlation is substantially lower for the group of the most traded bonds. Consistent with the finding in the time series analysis, panel B of Table 8 shows that the highest correlations are between CS-based estimators and *IQR* for most subsamples.[13] Generally, the correlations decrease from subsample 1 to subsample 4 and are higher when the negative spread estimations are discarded.

Finally, for each pair of measures and each month, I compute the difference (bias) in the mean values and in the three quartiles of the cross-sectional distribution of bonds within each subsample. The average biases for all months are reported in Table 9. Asterisks indicate rejection of the null that the average bias is zero by applying the Fama–MacBeth (1973) procedure. The main conclusions from the time series bias analysis in

Table 6 are repeated now. The lower downward biases in volatility estimates are shown for the adjusted versions.

Using *Roll* as the proxy for the true transaction cost, estimates that impose a zero spread when the spread is negative produce a non-significant bias for the median of the cross-sectional distribution in the subsample of highly traded bonds, with the lowest bias for Adjusted. In subsample 2, Non-zero and Jensen Non-zero approximate well the first quartile and the median, while Adjusted Non-zero and Adjusted Jensen Non-zero are better at reproducing the mean and Q3. For the other bonds, with a lower trading frequency, either Non-zero or Adjusted Jensen Non-zero is the best estimator, depending on the subsample and the statistic. Regarding *IQR*, the best estimator is Adjusted for all moments of the distribution in subsample 1. For the remaining bonds, measures that impose a zero spread in cases of negative estimates are downward biased. In subsample 2, Adjusted Non-zero for the mean and Q3, Non-zero for Q1, and Jensen Non-zero for Q2 are unbiased. For subsamples 3 and 4, Non-zero produces the lowest biases. Finally, in subsample 1, the four versions that impose a zero spread when it is negative are accurate in estimating the mean and Q3, while the other four are the best in terms of the median. For the other subsamples, Jensen Non-zero generally shows the lowest biases.

Summarizing the results from both the time series and cross-sectional analysis, on the one hand, the standard implementation of the CS estimator produces large, negative biases in the volatility component. The magnitude of this bias is directly related to bond volatility (inversely related to frequency of bond trading). The volatility estimation clearly improves when I use the adjusted version that accounts for gaps in trading and even slightly improves if Jensen's inequality is taking into account. On the other hand, regarding the spread component, the standard CS estimator is upward biased for highly traded bonds and downward biased if the average frequency of bond trading is lower than 1.0252 days. The adjustment for non-consecutive trading days reduces the spread estimates and then contributes to reducing the upward bias in subsample 1. However, both standard and adjusted estimators exhibit a problematic decreasing trend from the most liquid bonds to the least liquid bonds and thus negative and increasing biases from subsample 2 to subsample 4. The problem is considerably reduced if negative spreads are not imposed to equal zero. In such a case, the adjusted versions work reasonably well in subsample 2 and, depending on the intraday benchmark, in subsample 3. In addition, the Jensen Non-zero estimator seems to be the best option in subsample 4. For subsamples 2 to 4, the estimator adjusted by the number of days between trades without imposing a zero spread (with or without Jensen's inequality) shows biases that are considerably lower in absolute value than in standard practice for all three intraday benchmarks.

6.3. Biases in spread estimators for the whole sample

As said before, all the adjustments analyzed contribute to improving the estimation of the transaction cost under the CS model for the corporate bond market. However, the adjustments act in opposite directions and the adjustment that works the best depends on the bond's activity level. To arrive at a general conclusion, I repeat the bias analysis in both the time series and cross-sectionally, employing the whole sample of bonds in this case. Remember that the sample was selected with the standard requirements of most papers addressing liquidity issues in this market. Therefore, to obtain a general conclusion is relevant.

Following the same procedure described in previous sections, Table 10 displays biases in the time series distribution, in panel A, and biases in the cross-sectional distribution, in panel B. In this case, besides the standard CS estimator, I report results only for the three alternative estimators that can outperform it, given the previous findings.

Starting with the results shown in panel A, we see that the standard CS measure is downward biased compared with any of the intraday proxies. In the case of *Roll*, the bias is large, -0.14% (-0.21%) in the mean (median), is observed for the complete distribution, and is significantly different from zero. When negative spread cases are discarded (and not substituted by zero), the measure significantly overstates the transaction cost for the mean and the three quartiles. In this case, the bias increases from Q1 to Q3, indicating that the maximum error occurs during the crisis, when the transaction costs have the highest values. The adjustment for infrequent trading yields much smaller biases, from a minimum of 0.01% for the median to a maximum of 0.097% for the mean, both insignificantly different from zero. Again, the bias increases with the level of the spread and is significant for Q3 but its magnitude is less than one-half that for the Standard or Standard Non-zero measure. Incorporation of Jensen's inequality slightly increases the bias regarding the previous case generally, but it is the only proxy with a zero bias in Q1 (stability periods). The comparison with *IQR* again shows that the Standard CS proxy understates transaction costs. However, this measure now produces the lowest biases in absolute value for Q1 and the median. While the Adjusted Non-zero estimator shows the lowest bias in the case of Q3. Finally, in comparison with the round-trip cost measure, all the daily estimates are downward and significantly biased, but the Standard CS proxy produces biases that are more than double those of the other three proxies.

Figure 6 represents the time series of the each of the four CS spread estimators and the intraday proxies for the whole sample of bonds. Panel A shows how the Standard measure underestimates the spread. It is below the three intraday proxies in all moments, with the exception of the period between October 2008 and November 2009, when it exceeds the values of the *Roll* and *Round* measures. Panel B of Figure 6 illustrates the high upward bias that discarding negative spreads produces during recession periods in relation with all the intraday measures. Once the adjustment for infrequent trading is also incorporated, the previous overestimation of the spread during high-volatility periods is compensated for by the corresponding reduction in spread when volatility is not required to be zero. The result is that the final spread estimator is larger than the standard one in normal times and maintains reasonable values during the crisis (panel C of Figure 6). Finally, panel D of Figure 6 shows that Jensen's inequality produces a small and general increase in the spread in relation to the previous case.

Panel B of Table 10 displays biases in the cross-sectional distribution. In relation to the *Roll* proxy, the lowest biases are observed for the Adjusted Non-zero estimator, with insignificant biases of around 0.02% for both the mean and Q3. Incorporation of Jensen's inequality even improves the accuracy of the estimator for the median value. Regarding the *Round* proxy, all the CS-based estimators are downturn biased for the three quartiles but considerably larger biases arise for the standard measure. Lastly, compared with *IQR*, the Adjusted Non-zero estimator, with or without Jensen's inequality, generally produces the lowest bias and the bias is insignificant for both the mean and the median.

7. Summary and conclusion

Measuring liquidity in over-the-counter bond markets is not an easy task. This is the motivation behind SSU's paper that compares the performance of a very large set of liquidity proxies, including alternative measures for transaction costs, price impact, and price dispersion, computed from both high- and low-frequency data. SSU find that, among the measures that require only daily transaction data, the CS proxy precisely estimates the magnitude of the transaction costs for different bonds and their variation over time. The CS proposal uses only the high and low prices for each day and provides an estimator for daily volatility and for the daily bid–ask spread. SSU compute the CS measure following the standard methodology, which implies i) requiring that the high and low prices in a day without trades are the same as those observed in the previous transaction day, ii) setting a zero spread for days with a negative estimation, and iii) ignoring Jensen's inequality for computational efficiency.

Using daily transaction data in the period from October 1, 2004, to September 30, 2012, and bonds with at least one year of active trading and traded on at least 75% of the trading days, I observe that 16% of working days show no trades, on average, and that around 30% of daily CS transaction cost estimates are negative. Moreover, the infrequent trading is mainly concentrated during the crisis, when, additionally, volatility is higher and liquidity measurement becomes even more important. Therefore and independently of the appropriateness of the theoretical assumptions that support the CS model, the requirements of the measure's practical estimation can be problematic. The assumption of zero volatility when there are no observable trades will understate volatility and overstate the spread. In contrast, the imposition of a zero spread instead of excluding negative daily estimates will understate the spread. Finally, ignoring Jensen's inequality allows for closed-form solutions but could skew the estimation results.

The contribution of this paper is twofold. On the one hand, I propose an adjusted version of the CS estimator that does not require continuous trading. The basis of the model is the same as in the CS proposal but the estimator now combines information at a daily frequency with information at a lower frequency, which is defined by the observation of prices in the data. Thus, there is no estimation in days without trades and the estimator incorporates an adjustment when there is more than one day between the current transaction and the previous one. The results show that the adjusted version produces significantly higher values for volatility and lower values for the spread than for the standard version.

On the other hand, I evaluate the accuracy of standard and adjusted volatility and spread estimators in combination with the other practical estimation concerns, such as setting a zero spread when it is estimated to be negative, as opposed to deleting negative values, and ignoring Jensen's inequality, as opposed to incorporating it. To do so, I use realized volatility and three transaction cost measures based on high-frequency data as the true volatility and spread, respectively. The analysis is carried out by splitting the sample of bonds into four subsamples regarding the number of days without trades so that I can analyze the effects in relation to the bonds' level of liquidity (activity).

The empirical results show that the standard CS procedure underestimates the volatility for all the bonds and, more importantly, the volatility estimate decreases from the least volatile bonds to the most volatile bonds. The adjustment that accounts for gaps in trading clearly reduces the magnitude of this downward bias and reproduces the reasonable increasing trend in volatility as the frequency of trading diminishes. In addition, the incorporation of Jensen's inequality even slightly improves the accuracy of volatility estimates.

Regarding the spread, the standard CS procedure underestimates the transaction cost for the whole sample of bonds and the estimates decrease as the bond's illiquidity increases. The estimator shows a small but positive bias in subsample 1 (highly active bonds) driven by the imposition of zero volatility on days without trades, while a large, negative and increasing bias in subsamples 2 to 4 mainly driven, in this case, by the negative correlation between volatility and spread that the model imposes. The adjustment for non-consecutive trades contributes to reducing the upward bias found in subsample 1. However, standard and adjusted estimators are affected by the problem of high volatility when the frequency of trading decreases. The downward bias is considerably reduced if negative spreads are excluded. In this case, the adjusted estimator works reasonably well for the 50% of the bonds in the central position of the whole sample in terms of trading frequency. Finally, Jensen's inequality contributes to improving the accuracy of estimators in the cross section for all subsamples, but the effect is not clear in the times series dimension. In any case, differences regarding the corresponding versions that ignore Jensen's inequality are not significant and do not compensate for the increase in the computational cost that this numerical solution requires.

Overall, I can conclude that relaxing theoretical or practical impositions yields estimators that produce lower biases in terms of absolute value than the standard CS estimator for all subsamples and the three intraday benchmarks, in both time series and cross-sectionally. The spread estimator that accounts for gaps in trading and discards negative values does a good job.

Acknowledgments

This work was supported by the Generalitat Valenciana (PROMETEO/2017/158), the Spanish Department of Economy and Competitiveness (ECO2015-67035-P and ECO2014-58434-P), and the Bank of Spain (Excellence in Education and Research Grant 2016). I thank David Abad and Gonzalo Rubio for their unfailingly appropriate and helpful comments.

References

Amihud, Yakov, 2002. Illiquidity and stock returns: Cross-section and time-series effects. Journal of Financial Markets 5, 31–56. https://doi.org/10.1016/S1386-4181(01)00024-6.

Bao, Jack, Pan, Jun, Wang, Jiang, 2011. The illiquidity of corporate bonds. Journal of Finance 66 (3), 911–946. http://dx.doi.org/10.1111/j.1540-6261.2011.01655.x.

Bongaerts, Dion, de Jong, Frank, Driessen, Joost, 2017. An asset pricing approach to liquidity effects in corporate bond markets. Review of Financial Studies 30 (4), 1229–1269. https://doi.org/10.1093/rfs/hhx005.

Corwin, Shane A., Schultz, Paul, 2012. A simple way to estimate bid–ask spreads from daily high and low prices. Journal of Finance 67, 719–760. http://dx.doi.org/10.1111/j.1540-6261.2012.01729.x.

De Jong, Frank, Driessen, Joost, 2012. Liquidity risk premiums in corporate bond markets. Quarterly Journal of Finance 2 (2). https://doi.org/10.1142/S2010139212500061

Dick-Nielsen, Jens, 2009. Liquidity biases in TRACE. Journal of Fixed Income 19 (2), 43–55. https://doi.org/10.3905/jfi.2009.19.2.043.

________, 2014. How to clean enhanced TRACE data. SSRN working paper, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2337908.

Edwards, Amy K., Harris, Lawrence E., Piwowar, Michael S., 2007. Corporate bond market transaction costs and transparency. Journal of Finance 62 (3), 1421–1451. http://dx.doi.org/10.1111/j.1540-6261.2007.01240.x.

Fama, Eugene, MacBeth, James, 1973. Risk, return, and equilibrium: Empirical tests. Journal of Political Economy 81 (3), 607–636.

Feldhütter, Peter, 2012. The same bond at different prices: Identifying search frictions and selling pressure. Review of Financial Studies 25 (4), 1155–1206. DOI: https://doi.org/10.1093/rfs/hhr093.

Feller, William, 1951. The asymptotic distribution of the range of sums of independent random variables. Annals of Mathematical Statistics 22 (3), 427–432.

Friewald, Nils, Jankowitsch, Rainer, Subrahmanyam, Marti G., 2012. Illiquidity or credit deterioration: A study of liquidity in the U.S. corporate bond market during financial crises. Journal of Financial Economics 105 (1), 18–36. https://doi.org/10.1016/j.jfineco.2012.02.001.

Garman, Mark B., Klass, Michael J., 1980. On the estimation of security price volatilities from historical data. Journal of Business 53, 67–78.

Han, Song, Zhou, Hao, 2007. Nondefault bond spread and market trading liquidity. Working paper, Federal Reserve Board.

Parkinson, Michael, 1980. The extreme value method for estimating the variance of the rate of return. Journal of Business 53, 61–65.

Pu, Xiaoling, 2009. Liquidity commonality across the bond and CDS markets. Journal of Fixed Income 19 (1), 26-39. DOI: https://doi.org/10.3905/JFI.2009.19.1.026.

Roll, Richard, 1984. A simple implicit measure of the effective bid–ask spread in an efficient market. Journal of Finance 39 (4), 1127–1139. DOI: 10.2307/2327617.

Rossi, Marco, 2014. Realized volatility, liquidity, and corporate yield spreads. SSRN working paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1571437.

Schestag, Raphael, Schuster, Philipp, Uhrig-Homburg, Marliese, 2016. Measuring liquidity in bond markets. Review of Financial Studies 29 (5), 1170–1219. https://doi.org/10.1093/rfs/hhv132.

Table 1. Descriptive statistics of the TRACE database and selected bonds

Panel A: All TRACE data after filters for duplicates, cancellations, corrections, reversals, and outliers. Only market days and dates during the bond's life and until three months up to the default are included. The sample period is from July 1, 2002, to December 31, 2014.

Number of bonds:	89,654				
Industry distribution:	*Industrial*	*Financial*	*Utilities*	*Government*	*Miscellaneous*
	16.46%	56.09%	3.36%	23.95%	0.15%

	Mean	*St. Dev.*	*Q1*	*Q2*	*Q3*
Time to maturity at issuance	9.769	14.964	3.016	5.247	10.047
Offering amount (millions USD)	178.074	446.565	3.63	20	150
Coupon	3.643	2.972	0.95	3.54	5.65
Treasury spread at issuance	162.949	126.494	84	125	197
Rating (average in the bond's life)	5.397	4.220	1.25	5	7.833
Days with trade (%)	40.215	47.292	5.349	17.518	60.252
Average number of trades per day	4.113	11.555	2	2.560	3.778
Total number of trades per bond	971.399	4,285.111	11	61	352

Panel B: Selected bonds with liquidity requirements: One year between the first and last transaction dates and trades on at least 75% of trading days. The sample period is from October 1, 2004, to September 30, 2012.

Number of bonds:	3,526				
Industry distribution:	*Industrial*	*Financial*	*Utilities*	*Government*	*Miscellaneous*
	50.77%	40.36%	3.37%	5.39%	0.11%

	Mean	*St. Dev.*	*Q1*	*Q2*	*Q3*
Time to maturity at issuance	10.713	8.619	5.027	9.826	10.055
Offering amount (millions USD)	1,068.557	967.832	500	750	1,250
Coupon	5.539	2.093	4.350	5.625	6.875
Treasury spread at issuance	176.132	135.745	85	134	215
Rating (average in the bond's life)	7.610	4.076	4.988	6.750	9.891
Days with trade (%)	90.020	7.885	83.154	91.314	97.478
Average number of trades per day	10.893	9.869	5.234	7.583	12.452
Total number of trades per bond	10,107.984	11,878.617	3,355	6,677	12,259

This table reports the descriptive statistics of bond characteristics for the overall TRACE data set (panel A) and the selected sample (panel B). The Treasury spread is the difference between the bond yield and the yield on Treasury securities with the same maturity. The Treasury yields for all possible maturities are obtained by interpolation. Days with trades refer to the percentage of days with at least one trade in the bond's trading life span.

Table 2. Descriptive statistics of the subsamples of bonds, based on trading frequency

	Panel A: Subsample 1 (882 bonds) # Days between trades (average) < 1.0252					*Panel B: Subsample 2 (881 bonds) #Days between trades (average) in [1.0252, 1.0948)*				
Industry distribution:	*Industrial*	*Financial*	*Utilities*	*Govern.*	*Misc.*	*Industrial*	*Financial*	*Utilities*	*Govern.*	*Misc.*
	39.34%	52.15%	0.34%	8.16%	0.00%	52.21%	39.50%	2.95%	5.22%	0.11%
	Mean	*St. Dev.*	*Q1*	*Q2*	*Q3*	*Mean*	*St. Dev.*	*Q1*	*Q2*	*Q3*
Time to maturity at issuance	8.899	6.344	5.022	8.516	10.030	9.993	7.952	5.022	9.205	10.041
Offering amount (millions USD)	1,771.502	1,209.557	1,000	1,500	2,250	1,083.670	868.215	501.500	875	1,250
Coupon	5.156	1.684	4.125	5.3	6.125	5.466	2.151	4.250	5.500	6.804
Treasury spread at issuance	172.048	122.116	90	135	210	184.016	150.689	83	135	225
Rating (average in the bond's life)	6.125	3.457	4.333	5.712	7.623	7.395	4.083	4.667	6.456	9.500
Days with trade (%)	98.991	0.715	98.499	99.079	99.531	94.760	1.810	93.220	94.926	96.432
Average number of trades per day	21.300	13.601	12.353	17.565	24.994	9.668	4.772	6.826	8.641	10.946
Total number of trades per bond	19958.4	18210.7	7959	15618.5	25574	9311.1	6866.3	4380.3	7428	12714
Average number of days between trades	1.010	0.007	1.005	1.009	1.015	1.055	0.020	1.036	1.053	1.072
Maximum number of trades in a day	218.832	248.895	84	153	266	146.583	204.181	54	95	157
Average trading volume per day (000s USD)	7,254.8	6,251.8	3,458.1	5,577.2	8,921.6	3,816.6	3,073.9	1,972.1	3,091.5	4,707.4
Max. trading volume in a day (000s USD)	111,962.5	83,727.3	55,596.5	91,910.8	140,409.1	72,928.3	64,323.4	36,881.2	56,763.1	86,696.9
Volatility of daily returns (%)	1.154	0.802	0.568	1.006	1.483	1.464	1.174	0.668	1.176	1.890
	Panel C: Subsample 3 (882 bonds) #Days between trades (average) in [1.0948, 1.2015)					*Panel D: Subsample 4 (881 bonds) # Days between trades (average) ≥ 1.2015*				
Industry distribution:	*Industrial*	*Financial*	*Utilities*	*Govern.*	*Misc.*	*Industrial*	*Financial*	*Utilities*	*Govern.*	*Misc.*
	55.33%	35.49%	4.88%	4.08%	0.23%	56.19%	34.28%	5.33%	4.09%	0.11%
	Mean	*St. Dev.*	*Q1*	*Q2*	*Q3*	*Mean*	*St. Dev.*	*Q1*	*Q2*	*Q3*
Time to maturity at issuance	12.108	9.558	5.060	10	12.011	11.853	9.775	5.062	9.995	11.937
Offering mount (millions USD)	808.520	678.782	450	600	1,000	610.032	546.459	300	500	750
Coupon	5.708	2.184	4.400	5.750	7.125	5.826	2.244	4.7	6	7.2
Treasury spread at issuance	171.981	128.104	87	135	212.5	177.045	143.335	83	127	220
Rating (average in the bond's life)	8.314	4.109	5.667	7.891	10.583	8.657	4.153	5.667	8.278	11.613
Days with trade (%)	87.455	2.369	85.509	87.596	89.501	78.867	2.310	76.864	78.808	80.696
Average number of trades per day	6.970	3.925	4.946	5.988	7.644	5.629	3.675	3.901	4.675	6.076
Total number of trades per bond	6,492.8	5,378.0	2,662	5,200.5	8,721	4,662.6	4,052.9	1,914.8	3,552	6,349.5
Average number of days between trades	1.144	0.031	1.117	1.141	1.168	1.268	0.037	1.237	1.268	1.300
Maximum number of trades in a day	102.644	108.004	41	72	120	81.402	93.440	30	55	96
Average trading volume per day (000s USD)	2,861.4	2,346.0	1,481.9	2,311.0	3,545.3	2,144.3	1,746.7	984.4	1,825.8	2,769.5
Max. trading volume in a day (000s USD)	57,489.8	75,633.8	26,674.3	43,926.8	69,898.2	42,582.8	36,197.6	18,469.1	34,192.9	55,140.1
Volatility of daily returns (%)	1.686	1.245	0.840	1.332	2.211	1.735	1.305	0.846	1.390	2.187

Panels A to D report the descriptive statistics of security characteristics and activity indicators for the corporate bonds in each subsample. The sample selection criteria consist of bonds active for at least one year and traded on at least 75% of trading days. The subsamples are created by splitting the sample on the basis of the quartiles for the average number of days between trades, as indicated in each panel. Days with trades refer to the percentage of days with at least one trade in the bond's trading life. Volatility is the standard deviation of the daily returns from the closing prices for all days in the bond's life.

Table 3. Volatility and spread estimators: Mean values

Panel A: Volatilities				
	Subsample 1	Subsample 2	Subsample 3	Subsample 4
Standard	1.0450	1.0433	1.0296	0.9378
% zeros	2.15	9.40	19.15	29.09
Adjusted	1.0541	1.0873	1.1334	1.1093
% zeros	0.99	3.76	5.56	6.23
Diff. St-Adj.	-0.93	-4.36	-10.61*	-19.47*
Jensen	1.0513	1.0482	1.0792	0.9293
% zeros	2.15	9.40	19.16	29.09
Diff. St-Jensen	-0.70	-0.64	-0.11	0.62
Adjusted Jensen	1.0607	1.0932	1.1350	1.1043
% zeros	1.00	3.76	5.57	6.23
Diff. St-Adj. Jen	-1.66	-5.07	-10.92*	-19.21*
Panel B: Spreads				
	Subsample 1	Subsample 2	Subsample 3	Subsample 4
Standard	0.8898	0.7144	0.7508	0.7203
% zeros	19.28	28.85	32.51	32.40
Adjusted	0.8730	0.6532	0.5964	0.5147
Diff. St-Adj.	2.51	9.89*	21.13*	30.65*
Jensen	0.8969	0.7200	0.7520	0.7230
Diff. St-Jensen	-0.80	-0.65	-0.42	-0.27
St. Non-zero	1.0513	0.9872	1.0908	1.0459
Diff. St-NonZ	-26.07*	-44.47*	-50.65*	-49.85*
Adjusted Non-zero	1.0322	0.9187	0.9216	0.8440
Diff. St-Adj. NonZ	-23.07*	-32.98*	-27.44*	-19.25*
Jensen Non-zero	1.0616	0.9736	1.0439	1.0481
Diff. St-Jen NonZ	-28.07*	-45.75*	-53.15*	-52.12*
Adjusted Jensen	0.8804	0.6548	0.5829	0.5159
Diff. St-Adj. Jen	1.55	8.96*	20.15*	30.02*
Adj. Jensen Non-zero	1.0447	0.9161	0.9331	0.8614
Diff. St-Adj. Jen NonZ	-25.14*	-34.57*	-30.77*	-22.44*

This table reports the average across all bonds within each subsample of the mean value of daily volatility (spread) estimators in panel A (B), in percentages. The subsamples are based on the average number of days between trades and are described in Table 2. In panels A and B Standard (or St) refers to the standard practical application of the CS measure; Adjusted (or Adj.) refers to the estimator that accounts for infrequent trading; Jensen is the CS proposal that incorporates the Jensen's inequality; and Adjusted Jensen is the corresponding adjusted version. Additionally, in panel B, Non-zero (or NonZ), in the Standard, Adjusted, Jensen or Adjusted Jensen, refers to the practical approach of deleting negative spread cases. Diff. refers to the relative difference in percentage between the two indicated estimators. The superscript * indicates rejection of the null that the two compared estimators have the same mean. % zeros indicates the percentage of days, on average, that show a zero value for the volatility estimate, in panel A, and the percentage of days, on average, that the spread is negative and is then set to zero in the standard case, in panel B.

Table 4. Time series distribution of the aggregate measures for each subsample

Panel A: Volatilities

	Subsample 1					Subsample 2					Subsample 3					Subsample 4				
	Mean	*St.Dev*	*Q1*	*Q2*	*Q3*	*Mean*	*St.Dev*	*Q1*	*Q2*	*Q3*	*Mean*	*St.Dev*	*Q1*	*Q2*	*Q3*	*Mean*	*St.Dev*	*Q1*	*Q2*	*Q3*
Realized	1.204	0.616	0.845	0.938	1.293	1.446	0.717	1.020	1.208	1.432	1.566	0.760	1.107	1.314	1.553	1.590	0.757	1.123	1.364	1.608
Standard	1.062	0.417	0.826	0.924	1.046	1.117	0.515	0.803	0.955	1.100	1.126	0.514	0.822	0.966	1.127	1.065	0.482	0.773	0.904	1.091
Adjusted	1.074	0.431	0.833	0.931	1.071	1.170	0.562	0.836	0.986	1.151	1.258	0.629	0.896	1.041	1.255	1.285	0.660	0.895	1.081	1.277
Jensen	1.070	0.418	0.833	0.932	1.055	1.125	0.516	0.810	0.963	1.109	1.128	0.512	0.826	0.970	1.129	1.060	0.476	0.772	0.904	1.086
Adj. Jen	1.082	0.432	0.840	0.939	1.080	1.178	0.563	0.843	0.994	1.160	1.261	0.626	0.901	1.048	1.258	1.280	0.650	0.895	1.080	1.277

Panel B: Spreads

	Subsample 1					Subsample 2					Subsample 3					Subsample 4				
	Mean	*St.Dev*	*Q1*	*Q2*	*Q3*	*Mean*	*St.Dev*	*Q1*	*Q2*	*Q3*	*Mean*	*St.Dev*	*Q1*	*Q2*	*Q3*	*Mean*	*St.Dev*	*Q1*	*Q2*	*Q3*
Roll	0.910	0.307	0.728	0.826	0.971	1.054	0.306	0.856	0.986	1.154	1.182	0.323	0.969	1.103	1.315	1.221	0.419	0.992	1.137	1.369
IQR	0.794	0.477	0.520	0.599	0.902	1.005	0.526	0.704	0.844	1.078	1.129	0.523	0.817	0.973	1.211	1.161	0.505	0.851	1.030	1.249
Round	1.125	0.223	0.979	1.073	1.205	1.242	0.227	1.079	1.205	1.358	1.374	0.258	1.193	1.331	1.494	1.443	0.310	1.240	1.401	1.598
Standard	1.046	0.615	0.700	0.834	1.123	0.832	0.469	0.570	0.689	0.877	0.835	0.434	0.595	0.700	0.876	0.848	0.381	0.625	0.749	0.907
Adjusted	1.030	0.580	0.696	0.827	1.115	0.766	0.392	0.534	0.643	0.826	0.680	0.296	0.504	0.593	0.736	0.610	0.225	0.463	0.551	0.680
Jensen	1.055	0.620	0.705	0.841	1.131	0.840	0.476	0.574	0.693	0.883	0.840	0.438	0.598	0.702	0.881	0.851	0.384	0.627	0.751	0.910
Non-zero	1.252	0.676	0.877	1.026	1.328	1.135	0.571	0.810	0.965	1.193	1.213	0.581	0.888	1.029	1.273	1.235	0.526	0.930	1.094	1.306
Adj. NonZ	1.238	0.644	0.873	1.019	1.321	1.071	0.505	0.775	0.919	1.137	1.066	0.457	0.804	0.936	1.133	1.024	0.385	0.791	0.928	1.106
Jen. NonZ	1.268	0.682	0.890	1.039	1.340	1.157	0.579	0.830	0.985	1.216	1.235	0.588	0.909	1.050	1.292	1.254	0.530	0.945	1.114	1.321
Adj. Jen	1.040	0.586	0.701	0.834	1.125	0.774	0.400	0.538	0.647	0.834	0.687	0.300	0.508	0.597	0.743	0.616	0.231	0.466	0.556	0.686
Adj Jen NonZ	1.254	0.650	0.885	1.035	1.337	1.095	0.513	0.795	0.941	1.157	1.093	0.464	0.827	0.960	1.154	1.052	0.398	0.813	0.951	1.137

This table reports information about volatility measures in panel A and spread measures in panel B. All the measures are computed daily for each bond and are then aggregated across all bonds in each subsample. The subsamples are based on the average number of days between trades and are described in Table 2. This table provides the mean, standard deviation, and three quartiles of the time series distribution of the aggregate measures as percentages. In panel A, Realized refers to the realized volatility computed with daily returns in a window containing the last three months. The other four measures are different CS-based volatility estimators. In panel B, the first three rows refer to high-frequency transaction cost proxies computed daily with intraday prices. Here, Roll is the negative autocovariance in intraday returns, IQR is the interquartile range of trade prices within a day, and Round is the daily average of the round-trip cost computed with the high and low prices of each set of intraday round-trip trades. The other eight measures are different versions of the CS spread estimator. Table 3 describes the CS-based volatility and spread estimators.

Table 5. Time series correlations between the aggregate measures for each subsample

Panel A: Volatilities

	Subsample 1	Subsample 2	Subsample 3	Subsample 4
Standard	0.992	0.998	0.998	0.995
Adjusted	0.991	0.999	0.999	0.999
Jensen	0.992	0.998	0.997	0.994
Adj. Jen	0.991	0.999	0.999	0.999

Panel B: Spreads

	Subsample 1			Subsample 2			Subsample 3			Subsample 4		
	Roll	IQR	Round	Roll	IQR	Round	Roll	IQR	Round	Roll	IQR	Round
Standard	0.890	0.888	0.825	0.807	0.884	0.664	0.688	0.827	0.637	0.634	0.821	0.675
Adjusted	0.899	0.908	0.830	0.816	0.875	0.707	0.701	0.801	0.651	0.469	0.762	0.570
Jensen	0.892	0.891	0.825	0.811	0.888	0.665	0.691	0.831	0.638	0.635	0.824	0.675
Non-zero	0.907	0.910	0.816	0.849	0.918	0.678	0.745	0.895	0.666	0.627	0.887	0.651
Adj. NonZ	0.914	0.923	0.823	0.835	0.893	0.702	0.728	0.822	0.633	0.492	0.819	0.574
Jen. NonZ	0.909	0.913	0.816	0.853	0.921	0.679	0.750	0.899	0.669	0.629	0.891	0.652
Adj. Jen	0.900	0.910	0.830	0.819	0.879	0.708	0.704	0.805	0.654	0.574	0.755	0.641
Adj Jen NZ	0.915	0.924	0.823	0.839	0.895	0.703	0.731	0.826	0.637	0.593	0.808	0.639

Panel A contains the time series correlations between four volatility CS-based estimators and realized volatility computed with daily returns in a window containing the last three months. Panel B displays the time series correlation between eight CS-based spread estimators and three high-frequency spread proxies computed daily with intraday prices. Table 3 describes the CS-based volatility and spread estimators. The high-frequency proxies are Roll, which is the negative autocovariance in intraday returns; IQR, which is the interquartile range of trade prices within a day; and Round, which is the daily average of the round-trip cost computed with the high and low prices of each set of intraday round-trip trades. All the measures are computed daily for each bond and are then aggregated among all bonds within each subsample. The subsamples are based on the average number of days between trades and are described in Table 2.

Table 6. Biases in the moments of the time series distribution of the aggregate measures for each subsample

Panel A: Volatilities																
	Subsample 1				Subsample 2				Subsample 3				Subsample 4			
	Mean	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*
Standard	-0.142*	-0.019*	-0.014	-0.247*	-0.329*	-0.217*	-0.252*	-0.332*	-0.440*	-0.285*	-0.348*	-0.427*	-0.526*	-0.350*	-0.460*	-0.517*
Adjusted	-0.130*	-0.012*	-0.007	-0.222*	-0.276*	-0.184*	-0.222*	-0.281*	-0.308*	-0.211*	-0.273*	-0.299*	-0.305*	-0.227*	-0.283*	-0.331*
Jensen	-0.134*	-0.013*	-0.006	-0.239*	-0.321*	-0.210*	-0.245*	-0.323*	-0.437*	-0.280*	-0.344*	-0.424*	-0.531*	-0.352*	-0.461*	-0.522*
Adj. Jen	-0.122*	-0.005*	0.000	-0.213*	-0.268*	-0.177*	-0.214*	-0.271*	-0.305*	-0.205*	-0.266*	-0.295*	-0.310*	-0.228*	-0.284*	-0.331*
Panel B: Spreads																
	Subsample 1				Subsample 2				Subsample 3				Subsample 4			
Roll	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*
Standard	0.133	-0.028*	0.007	0.152*	-0.225*	-0.286*	-0.297*	-0.277*	-0.350*	-0.374*	-0.404*	-0.440*	-0.375*	-0.368*	-0.389*	-0.463*
Adjusted	0.119	-0.032*	8E-04	0.144*	-0.289	-0.322*	-0.343*	-0.329*	-0.504*	-0.465*	-0.511*	-0.580*	-0.613*	-0.530*	-0.586*	-0.691*
Jensen	0.142*	-0.022*	0.015	0.159*	-0.217*	-0.283*	-0.293*	-0.271*	-0.345*	-0.371*	-0.401*	-0.434*	-0.372*	-0.366*	-0.386*	-0.461*
Non-zero	0.339*	0.149*	0.199*	0.355*	0.078*	-0.046*	-0.021*	0.039*	0.028*	-0.081*	-0.074*	-0.042*	0.013*	-0.063*	-0.044*	-0.064*
Adj. NonZ	0.326*	0.145*	0.192*	0.350*	0.016*	-0.081*	-0.067*	-0.017	-0.120*	-0.165*	-0.167*	-0.182*	-0.199*	-0.201*	-0.210*	-0.265*
Jen. NonZ	0.355*	0.162*	0.212*	0.369*	0.100	-0.027*	-0.001	0.061*	0.050*	-0.060*	-0.053*	-0.023	0.032*	-0.047*	-0.024*	-0.049*
Adj. Jen	0.128	-0.026*	0.008	0.154*	-0.281*	-0.318*	-0.339*	-0.321*	-0.498*	-0.461*	-0.506*	-0.573*	-0.607*	-0.526*	-0.581*	-0.684*
Adj Jen NZ	0.342*	0.158*	0.208	0.365*	0.039*	-0.062*	-0.046	0.002	-0.093	-0.142*	-0.143*	-0.161*	-0.171*	-0.179	-0.187*	-0.235*
IQR	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*
Standard	0.252*	0.180*	0.235*	0.221*	-0.173*	-0.134*	-0.155*	-0.201*	-0.294*	-0.222*	-0.273*	-0.335*	-0.314*	-0.227*	-0.281*	-0.342*
Adjusted	0.236*	0.176*	0.228*	0.214*	-0.239*	-0.170*	-0.201*	-0.252*	-0.449*	-0.313*	-0.380*	-0.476*	-0.552*	-0.389*	-0.478*	-0.570*
Jensen	0.261*	0.185*	0.242*	0.230*	-0.165*	-0.130*	-0.151*	-0.195*	-0.289*	-0.219*	-0.271*	-0.330*	-0.310*	-0.225*	-0.278*	-0.339*
Non-zero	0.458*	0.357*	0.427*	0.426*	0.129*	0.106*	0.121*	0.115*	0.084*	0.071*	0.056*	0.062*	0.074*	0.078*	0.064*	0.057*
Adj. NonZ	0.444*	0.353*	0.420*	0.420*	0.066*	0.071*	0.075*	0.059*	-0.063*	-0.013*	-0.037*	-0.078*	-0.137*	-0.060*	-0.102*	-0.144*
Jen. NonZ	0.474*	0.370*	0.440*	0.438*	0.152*	0.126*	0.141*	0.138*	0.106*	0.092*	0.077*	0.081*	0.092*	0.094*	0.084*	0.072*
Adj. Jen	0.246*	0.182*	0.235*	0.224*	-0.231*	-0.166*	-0.197*	-0.244*	-0.443*	-0.310*	-0.376*	-0.468*	-0.546*	-0.385*	-0.474*	-0.564*
Adj Jen NZ	0.46*	0.366*	0.436*	0.435*	0.090*	0.091	0.097*	0.079	-0.036	0.01	-0.012	-0.057*	-0.110*	-0.038*	-0.079*	-0.112*
Round	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*
Standard	-0.081*	-0.279*	-0.239*	-0.082*	-0.411*	-0.509*	-0.516*	-0.482*	-0.539*	-0.599*	-0.631*	-0.619*	-0.595*	-0.616*	-0.652*	-0.691*
Adjusted	-0.096*	-0.283*	-0.246*	-0.089*	-0.477*	-0.545*	-0.562*	-0.532*	-0.694*	-0.689*	-0.738*	-0.759*	-0.834*	-0.778*	-0.850*	-0.918*
Jensen	-0.072*	-0.274*	-0.232*	-0.074*	-0.404*	-0.505*	-0.512*	-0.476*	-0.534*	-0.595*	-0.629*	-0.613*	-0.592*	-0.614*	-0.650*	-0.688*
Non-zero	0.125*	-0.102*	-0.047*	0.122*	-0.109*	-0.269*	-0.240*	-0.165*	-0.161*	-0.306*	-0.302*	-0.222*	-0.208*	-0.311*	-0.308*	-0.292*
Adj. NonZ	0.112*	-0.106*	-0.054*	0.116*	-0.172*	-0.303*	-0.286*	-0.221*	-0.308*	-0.389*	-0.395*	-0.361*	-0.419*	-0.449*	-0.474*	-0.493*
Jen. NonZ	0.141*	-0.089*	-0.034*	0.135*	-0.087*	-0.249*	-0.220*	-0.143*	-0.139*	-0.284*	-0.281*	-0.203*	-0.189*	-0.295*	-0.287*	-0.277*
Adj. Jen	-0.086*	-0.278*	-0.239*	-0.080*	-0.469*	-0.541*	-0.558*	-0.525*	-0.687*	-0.686*	-0.734*	-0.752*	-0.828*	-0.774*	-0.846*	-0.912*
Adj Jen NZ	0.128*	-0.094*	-0.038*	0.132*	-0.148*	-0.284*	-0.264*	-0.203*	-0.281*	-0.366*	-0.370*	-0.340*	-0.391*	-0.427*	-0.450*	-0.461*

This table reports the differences in descriptive statistics (the mean and three quartiles) of the time series distribution between CS-based estimators and a benchmark. Panel A refers to volatility estimators and the benchmark measure is realized volatility computed with daily returns in a window containing the last three months. Panel B refers to spread estimators and the benchmark measure is one of the three high-frequency proxies computed daily with intraday prices: Roll, which is the negative autocovariance in intraday returns; IQR, which is the interquartile range of trade prices within a day; and Round, which is the daily average of the round-trip cost computed with the high and low prices of each set of intraday round-trip trades. Table 3 describes the CS-based volatility and spread estimators. All the measures are computed daily for each bond and are then aggregated among all bonds within each subsample. The subsamples are based on the average number of days between trades and are described in Table 2. The superscript * indicates that the null of equal means or equal quartiles is rejected.

Table 7. Distribution of the mean values of cross-sectional statistics

Panel A: Volatilities

	Subsample 1					Subsample 2					Subsample 3					Subsample 4				
	Mean	*St.Dev*	*Q1*	*Q2*	*Q3*	*Mean*	*St.Dev*	*Q1*	*Q2*	*Q3*	*Mean*	*St.Dev*	*Q1*	*Q2*	*Q3*	*Mean*	*St.Dev*	*Q1*	*Q2*	*Q3*
Realized	1.155	0.780	0.611	0.966	1.483	1.388	1.068	0.645	1.075	1.792	1.512	1.105	0.725	1.214	1.990	1.517	1.125	0.713	1.208	2.008
Standard	1.041	0.619	0.637	0.904	1.274	1.089	0.779	0.566	0.883	1.359	1.108	0.777	0.570	0.911	1.431	1.037	0.752	0.514	0.844	1.349
Adjusted	1.054	0.628	0.645	0.915	1.288	1.144	0.822	0.594	0.922	1.422	1.240	0.872	0.642	1.011	1.593	1.251	0.894	0.632	1.024	1.610
Jensen	1.049	0.622	0.642	0.912	1.283	1.097	0.782	0.571	0.890	1.369	1.109	0.776	0.572	0.914	1.434	1.032	0.750	0.512	0.839	1.341
Adj. Jen	1.061	0.630	0.650	0.923	1.298	1.152	0.826	0.598	0.929	1.433	1.242	0.871	0.645	1.013	1.598	1.246	0.888	0.632	1.020	1.603

Panel B: Spreads

	Subsample 1					Subsample 2					Subsample 3					Subsample 4				
	Mean	*St.Dev*	*Q1*	*Q2*	*Q3*	*Mean*	*St.Dev*	*Q1*	*Q2*	*Q3*	*Mean*	*St.Dev*	*Q1*	*Q2*	*Q3*	*Mean*	*St.Dev*	*Q1*	*Q2*	*Q3*
Roll	0.898	0.588	0.479	0.760	1.167	0.999	0.795	0.456	0.770	1.296	1.074	0.933	0.442	0.819	1.429	1.114	1.043	0.402	0.823	1.521
IQR	0.791	0.598	0.398	0.631	0.994	0.990	0.825	0.431	0.720	1.277	1.091	0.882	0.486	0.826	1.426	1.107	0.910	0.480	0.839	1.472
Round	1.090	0.764	0.632	0.961	1.392	1.158	0.808	0.592	0.955	1.504	1.280	1.278	0.597	1.030	1.668	1.331	1.442	0.578	1.059	1.776
Standard	1.041	0.849	0.452	0.801	1.371	0.828	0.791	0.304	0.558	1.066	0.832	0.779	0.308	0.583	1.090	0.843	0.768	0.318	0.609	1.127
Adjusted	1.020	0.844	0.435	0.780	1.349	0.748	0.753	0.255	0.488	0.959	0.650	0.675	0.202	0.420	0.854	0.569	0.603	0.166	0.366	0.767
Jensen	1.048	0.856	0.455	0.808	1.381	0.835	0.801	0.306	0.562	1.074	0.836	0.784	0.309	0.585	1.095	0.846	0.771	0.318	0.610	1.132
Non-zero	1.194	0.894	0.571	0.964	1.562	1.078	0.935	0.444	0.781	1.401	1.155	0.967	0.487	0.873	1.529	1.186	0.980	0.506	0.910	1.581
Adj. NonZ	1.171	0.887	0.554	0.944	1.533	0.986	0.883	0.390	0.697	1.282	0.955	0.852	0.367	0.691	1.279	0.905	0.828	0.338	0.662	1.216
Jen. NonZ	1.208	0.900	0.581	0.976	1.578	1.095	0.945	0.453	0.794	1.423	1.171	0.975	0.498	0.887	1.548	1.201	0.988	0.516	0.925	1.600
Adj. Jen	1.030	0.850	0.440	0.788	1.361	0.756	0.763	0.257	0.493	0.968	0.656	0.681	0.204	0.424	0.863	0.574	0.609	0.167	0.370	0.773
Adj Jen NonZ	1.185	0.893	0.563	0.956	1.550	1.004	0.894	0.402	0.714	1.303	0.975	0.866	0.379	0.710	1.305	0.927	0.846	0.350	0.680	1.244

This table reports information about volatility measures in panel A and spread measures in panel B. All the measures are computed daily for each bond and daily series are averaged monthly within all days in the month. Then, for each month, the mean, standard deviation, and three quartiles of the cross section of bonds in each subsample are computed. The table reports the average values of these statistics. The subsamples are based on the average number of days between trades and are described in Table 2. In panel A, Realized refers to realized volatility computed with daily returns in a window containing the last three months. The other four measures are different CS-based volatility estimators. The first three rows of panel B refer to high-frequency spread proxies computed daily with intraday prices: Roll, which is the negative autocovariance in intraday returns; IQR, which is the interquartile range of trade prices within a day; and Round, which is the daily average of the round-trip cost computed with the high and low prices of each set of intraday round-trip trades. The remaining rows refer to different versions of the CS spread estimator. Table 3 describes the CS-based volatility and spread estimators.

Table 8. Average of the cross-sectional correlations

Panel A: Volatilities

	Subsample 1	Subsample 2	Subsample 3	Subsample 4
Standard	0.712	0.852	0.857	0.859
Adjusted	0.714	0.855	0.865	0.872
Jensen	0.714	0.852	0.857	0.860
Adj. Jen	0.715	0.856	0.866	0.875

Panel B: Spreads

	Subsample 1			Subsample 2			Subsample 3			Subsample 4		
	Roll	IQR	Round	Roll	IQR	Round	Roll	IQR	Round	Roll	IQR	Round
Standard	0.803	0.753	0.722	0.778	0.819	0.717	0.679	0.781	0.651	0.609	0.753	0.623
Adjusted	0.797	0.745	0.717	0.758	0.790	0.694	0.648	0.735	0.615	0.569	0.686	0.574
Jensen	0.803	0.754	0.722	0.780	0.821	0.718	0.680	0.783	0.652	0.611	0.756	0.624
Non-zero	0.836	0.798	0.746	0.801	0.863	0.733	0.688	0.815	0.655	0.620	0.782	0.626
Adj. NonZ	0.832	0.792	0.743	0.792	0.848	0.721	0.674	0.788	0.633	0.594	0.731	0.588
Jen. NonZ	0.837	0.799	0.747	0.801	0.864	0.733	0.689	0.817	0.655	0.621	0.783	0.625
Adj. Jen	0.798	0.746	0.718	0.760	0.793	0.695	0.649	0.737	0.616	0.572	0.690	0.576
Adj Jen NZ	0.833	0.793	0.743	0.794	0.849	0.722	0.674	0.790	0.632	0.594	0.732	0.585

This table reports cross-sectional correlations between four CS-based volatility estimators and realized volatility in panel A and between eight CS-based spread estimators and three high-frequency proxies in panel B. Realized volatility is computed with daily returns in a window containing the last three months. The high-frequency spread proxies are computed daily with intraday prices. Roll is the negative autocovariance in intraday returns, IQR is the interquartile range of trade prices within a day and Round is the daily average of the round-trip cost computed with the high and low prices of each set of intraday round-trip trades. Table 3 describes the CS-based volatility and spread estimators. All the measures are computed daily for each bond and daily series are averaged monthly within all days in the month. Then, for each month and each pair of measures, the correlation of the cross section of bonds in each subsample is computed. The table reports the average value. The subsamples are based on the average number of days between trades and are described in Table 2.

Table 9. Average biases in cross-sectional moments

	Subsample 1				Subsample 2				Subsample 3				Subsample 4			
Panel A: Volatilities																
	Mean	Q1	Q2	Q3	Mean	Q1	Q2	Q3	Mean	Q1	Q2	Q3	Mean	Q1	Q2	Q3
Standard	-0.113*	-0.166*	0.027*	-0.060*	-0.209*	-0.296*	-0.291*	-0.077*	-0.190*	-0.430*	-0.397*	-0.327*	-0.151*	-0.297*	-0.550*	-0.471*
Adjusted	-0.101*	-0.157*	0.035*	-0.050*	-0.195*	-0.242*	-0.247*	-0.049*	-0.152*	-0.369*	-0.268*	-0.236*	-0.080*	-0.198*	-0.394*	-0.261*
Jensen	-0.106*	-0.163*	0.032*	-0.053*	-0.199*	-0.288*	-0.287*	-0.072*	-0.184*	-0.420*	-0.395*	-0.326*	-0.149*	-0.294*	-0.548*	-0.476*
Adj. Jen	-0.093*	-0.154*	0.040*	-0.042*	-0.185*	-0.234*	-0.244*	-0.044*	-0.144*	-0.358*	-0.266*	-0.236*	-0.077*	-0.197*	-0.390*	-0.266*
Panel B: Spreads																
	Subsample 1				Subsample 2				Subsample 3				Subsample 4			
Roll	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*
Standard	0.144*	-0.026	0.043	0.205*	-0.168*	-0.150*	-0.210*	-0.227*	-0.235*	-0.130*	-0.229*	-0.329*	-0.249*	-0.070*	-0.188*	-0.356*
Adjusted	0.123*	-0.043*	0.021	0.184*	-0.246*	-0.199*	-0.280*	-0.331*	-0.409*	-0.231*	-0.386*	-0.552*	-0.510*	-0.217*	-0.425*	-0.704*
Jensen	0.152*	-0.023	0.050	0.215*	-0.161*	-0.148*	-0.205*	-0.219*	-0.231*	-0.129*	-0.227*	-0.322*	-0.245*	-0.069*	-0.186*	-0.351*
Non-zero	0.298*	0.094*	0.205*	0.395*	0.081*	-0.010	0.013	0.109*	0.089*	0.051*	0.062*	0.114*	0.104*	0.127*	0.123*	0.114*
Adj. NonZ	0.275*	0.076*	0.186*	0.367*	-0.009	-0.063*	-0.071*	-0.011	-0.105*	-0.067*	-0.113*	-0.130*	-0.173*	-0.043*	-0.126*	-0.252*
Jen. NonZ	0.311*	0.103*	0.217*	0.412*	0.098*	-0.001	0.027	0.128*	0.105*	0.062*	0.077*	0.132*	0.119*	0.138*	0.137*	0.132*
Adj. Jen	0.133*	-0.038	0.029	0.196*	-0.238*	-0.196*	-0.275*	-0.322*	-0.403*	-0.229*	-0.382*	-0.544*	-0.504*	-0.215*	-0.421*	-0.697*
Adj Jen NZ	0.288*	0.085*	0.197*	0.384*	0.008	-0.052*	-0.054*	0.010	-0.086*	-0.054*	-0.096*	-0.106*	-0.151*	-0.031*	-0.107*	-0.224*
IQR	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*
Standard	0.250*	0.055*	0.171*	0.378*	-0.161*	-0.126*	-0.161*	-0.210*	-0.259*	-0.178*	-0.242*	-0.336*	-0.263*	-0.158*	-0.226*	-0.342*
Adjusted	0.229*	0.037*	0.149*	0.356*	-0.241*	-0.176*	-0.232*	-0.317*	-0.440*	-0.283*	-0.405*	-0.570*	-0.536*	-0.311*	-0.470*	-0.703*
Jensen	0.258*	0.058*	0.177*	0.388*	-0.154*	-0.124*	-0.157*	-0.202*	-0.254*	-0.176*	-0.240*	-0.330*	-0.259*	-0.157*	-0.224*	-0.337*
Non-zero	0.404*	0.173*	0.333*	0.568*	0.089*	0.014	0.062*	0.126*	0.064*	0.002*	0.047*	0.103*	0.083*	0.031*	0.077*	0.115*
Adj. NonZ	0.380*	0.156*	0.313*	0.539*	-0.003	-0.041*	-0.022*	0.005	-0.136*	-0.119*	-0.135*	-0.147*	-0.202*	-0.142*	-0.176*	-0.255*
Jen. NonZ	0.417*	0.183*	0.345*	0.585*	0.105*	0.023*	0.075*	0.147*	0.081*	0.013*	0.062*	0.123*	0.097*	0.042*	0.091*	0.132*
Adj. Jen	0.239*	0.042*	0.157*	0.368*	-0.233*	-0.173*	-0.227*	-0.308*	-0.434*	-0.281*	-0.401*	-0.562*	-0.531*	-0.310*	-0.467*	-0.696*
Adj Jen NZ	0.394*	0.165*	0.325*	0.556*	0.014	-0.029*	-0.006	0.026*	-0.116*	-0.107*	-0.116*	-0.121*	-0.180*	-0.130*	-0.158*	-0.227*
Round	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*
Standard	-0.031	-0.177*	-0.156*	-0.014	-0.317*	-0.283*	-0.390*	-0.425*	-0.404*	-0.282*	-0.435*	-0.555*	-0.441*	-0.253*	-0.443*	-0.635*
Adjusted	-0.052	-0.195*	-0.178*	-0.036	-0.397*	-0.333*	-0.461*	-0.532*	-0.585*	-0.388*	-0.597*	-0.789*	-0.714*	-0.406*	-0.686*	-0.996*
Jensen	-0.023	-0.174*	-0.149*	-0.004	-0.310*	-0.281*	-0.385*	-0.418*	-0.400*	-0.280*	-0.432*	-0.548*	-0.437*	-0.251*	-0.440*	-0.631*
Non-zero	0.122*	-0.058	0.007	0.176*	-0.067	-0.143*	-0.167*	-0.089	-0.081*	-0.102*	-0.144*	-0.115*	-0.094*	-0.062*	-0.137*	-0.177*
Adj. NonZ	0.099*	-0.076*	-0.014	0.148*	-0.159*	-0.198*	-0.251*	-0.210*	-0.281*	-0.223*	-0.326*	-0.365*	-0.378*	-0.234*	-0.390*	-0.545*
Jen. NonZ	0.136*	-0.049	0.018	0.193*	-0.051	-0.134*	-0.153*	-0.068*	-0.064	-0.091*	-0.130*	-0.096*	-0.080*	-0.051*	-0.123*	-0.160*
Adj. Jen	-0.042	-0.190*	-0.170*	-0.024	-0.389*	-0.330*	-0.456*	-0.523*	-0.579*	-0.386*	-0.593*	-0.781*	-0.709*	-0.404*	-0.682*	-0.989*
Adj Jen NZ	0.113*	-0.066*	-0.002	0.164*	-0.142*	-0.186*	-0.235*	-0.189*	-0.261*	-0.211*	-0.308*	-0.339*	-0.357*	-0.223*	-0.372*	-0.517*

This table reports the differences in descriptive statistics (the mean and three quartiles) of the cross-sectional distribution between four different CS volatility estimators and realized volatility in panel A and between eight CS-based spread estimators and three high-frequency proxies in panel B.

Realized volatility is computed with daily returns in a window containing the last three months. The high-frequency spread proxies are computed daily with intraday prices. Roll is the negative autocovariance in intraday returns, IQR is the interquartile range of trade prices within a day and Round is the daily average of the round-trip cost computed with the high and low prices of each set of intraday round-trip trades. Table 3 describes the CS-based volatility and spread estimators. All the measures are computed daily for each bond and daily series are averaged monthly within all days in the month. Then, for each month and each pair of measures, the differences (biases) between the cross-sectional descriptive statistics are computed. The table reports the average values across the bonds in each subsample and * indicates that the null of equal means or equal quartiles is rejected. The subsamples are based on the average number of days between trades and are described in Table 2.

Table 10. Biases in spread estimators: Whole sample of bonds

Panel A: Biases in the time series distribution

	Roll				IQR				Round			
	Mean	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*
Standard	-0.137*	-0.217*	-0.213*	-0.173*	-0.091*	-0.076*	-0.081*	-0.135*	-0.366*	-0.486*	-0.473*	-0.418*
Non-zero	0.206*	0.043*	0.082*	0.167*	0.251*	0.185*	0.214*	0.205*	-0.023*	-0.225*	-0.178*	-0.077*
Adj. NonZ	0.097	-0.022*	0.010	0.081*	0.142*	0.119*	0.143*	0.120*	-0.132*	-0.291*	-0.249*	-0.163*
Adj Jen NZ	0.111*	-0.002	0.025*	0.096*	0.157*	0.139*	0.157*	0.134*	-0.118*	-0.271*	-0.235*	-0.149*

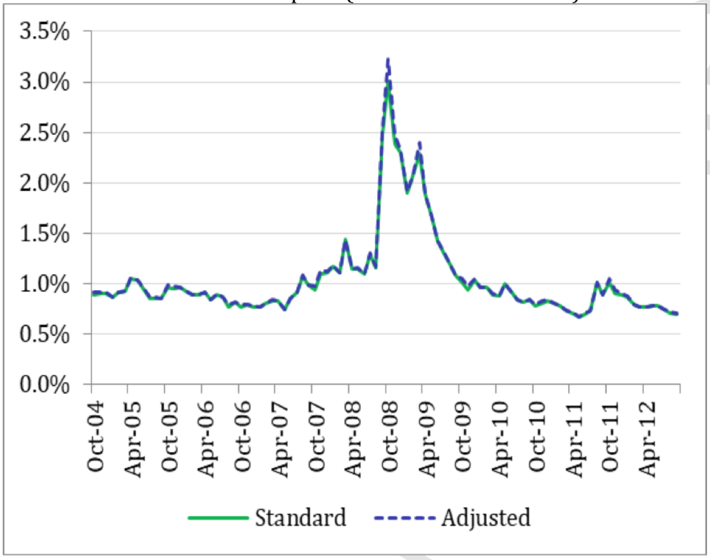
Panel B: Biases in the cross-sectional distribution

	Roll				IQR				Round			
	Mean	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*	*Mean*	*Q1*	*Q2*	*Q3*
Standard	-0.111*	-0.111*	-0.149*	-0.149*	-0.100*	-0.110*	-0.113*	-0.120*	-0.287*	-0.263*	-0.360*	-0.388*
Non-zero	0.183*	0.053*	0.102*	0.206*	0.193*	0.053*	0.138*	0.236*	0.006	-0.099*	-0.108*	-0.032
Adj. NonZ	0.025	-0.041*	-0.030*	0.021	0.027*	-0.047*	-0.002	0.041*	-0.159*	-0.199*	-0.248*	-0.227*
Adj Jen NZ	0.033*	-0.029*	-0.014	0.043	0.037*	-0.035*	0.014	0.064*	-0.150*	-0.187*	-0.232*	-0.204*

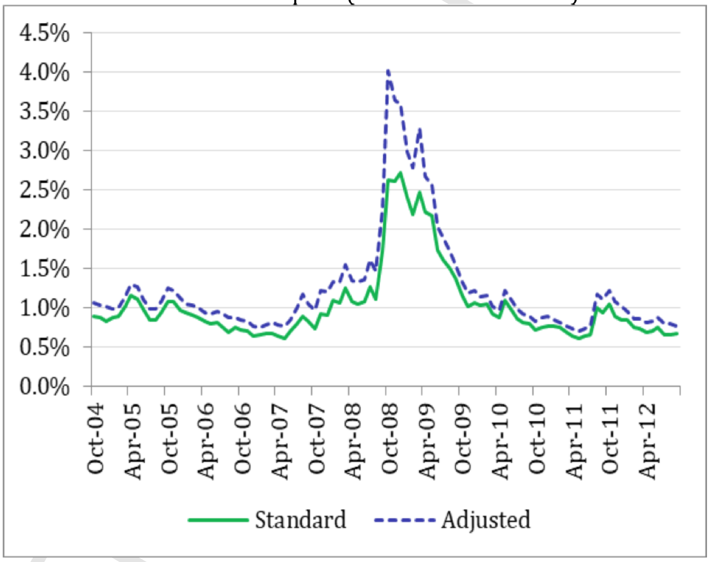
This table reports the differences in descriptive statistics (the mean and three quartiles) of the time series distribution in panel A and the cross-sectional distribution in panel B, between four different CS-based transaction cost estimators and three benchmarks: Roll, which is the negative autocovariance in intraday returns; IQR, which is the interquartile range of trade prices within a day; and Round, which is the daily average of the round-trip cost computed with the high and low prices of each set of intraday round-trip trades. The superscript * indicates that the null of equal means or equal quartiles is rejected. See notes in Tables 6 and 9 for details about the procedure to compute time series and cross-sectional biases, respectively.

Figure 1. Time series of the aggregate volatility estimators for the two extreme subsamples of bonds based on the average number of days between trades.

A. Subsample 1 (the most active bonds)



B. Subsample 4 (the least active bonds)



This figure displays the monthly time series of two volatility estimators: the CS proposal (Standard) and an adjusted version that accounts for infrequent trading (Adjusted). The two proxies are computed daily using the high and low prices from intraday TRACE data for the period between October 1, 2004, and September 30, 2012. The monthly values are averaged over all days within the month. The top (bottom) graph represents the average measures across all bonds in subsample 1 (4). The subsamples are based on the average number of days between trades and are described in Table 2.

Figure 2. Time series of the aggregate spread estimators across the bonds in subsample 1 (the most active bonds)

See the notes in Table 3 for the estimators' definitions.

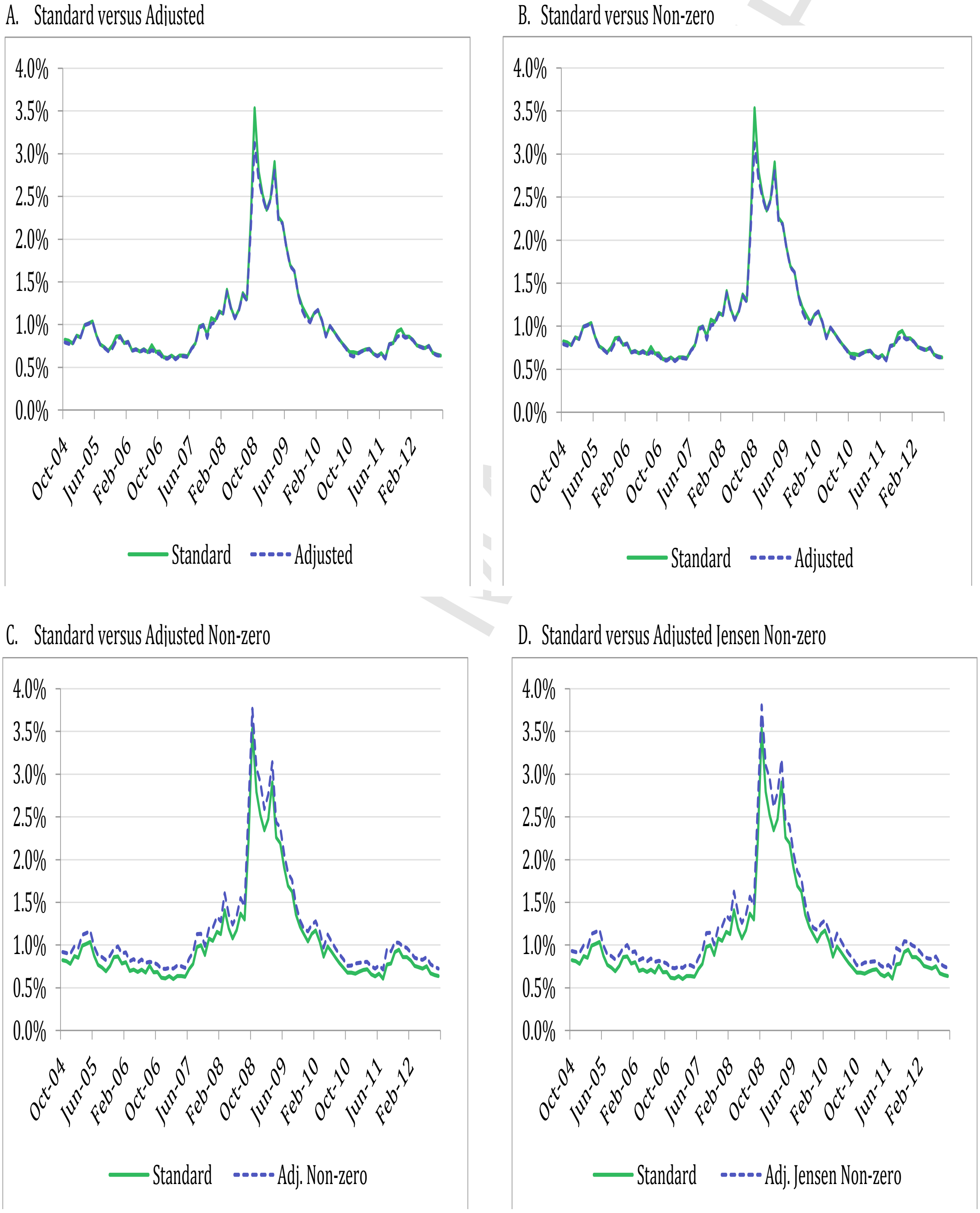
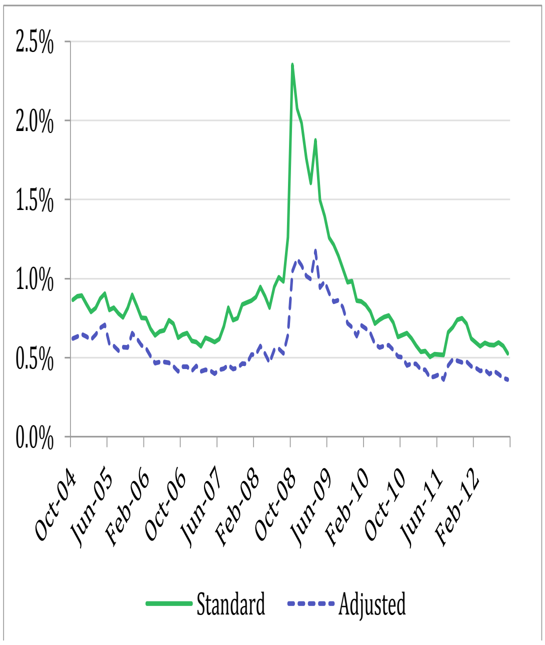
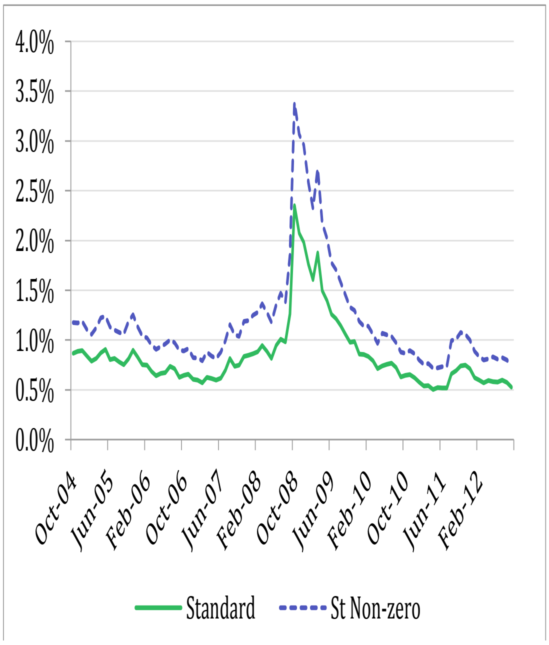


Figure 3. Time series of the aggregate spread estimators across the bonds in subsample 4 (the least active bonds).
See the notes in Table 3 for the estimators' definitions.

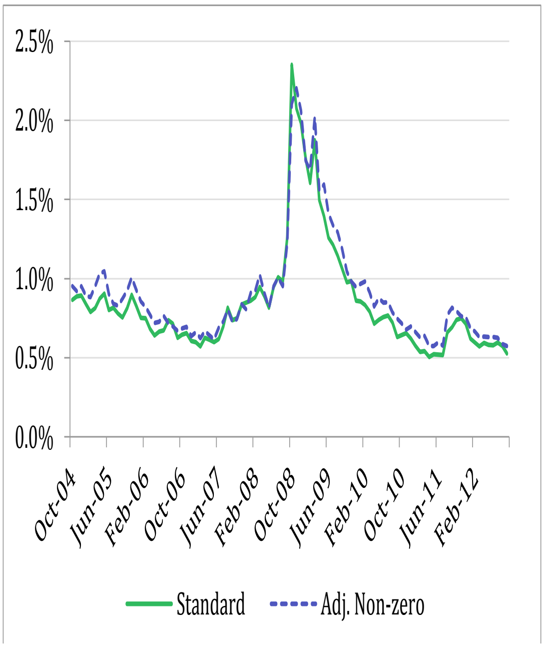
A. Standard versus Adjusted



B. Standard versus Non-zero



C. Standard versus Adjusted Non-zero



D. Standard versus Adjusted Jensen Non-zero

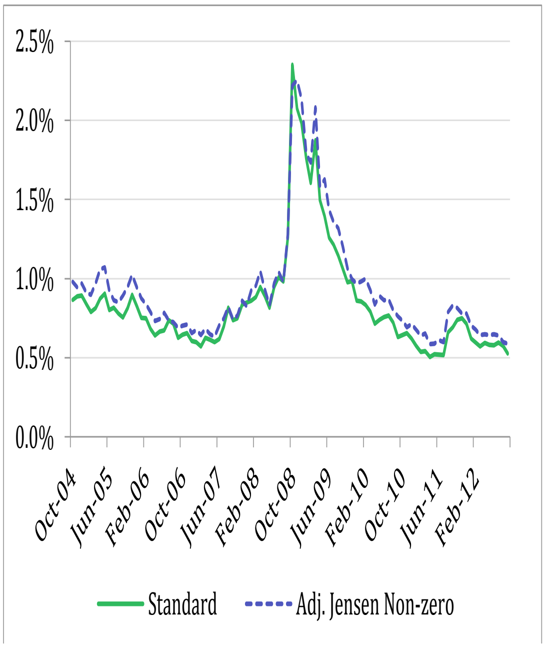
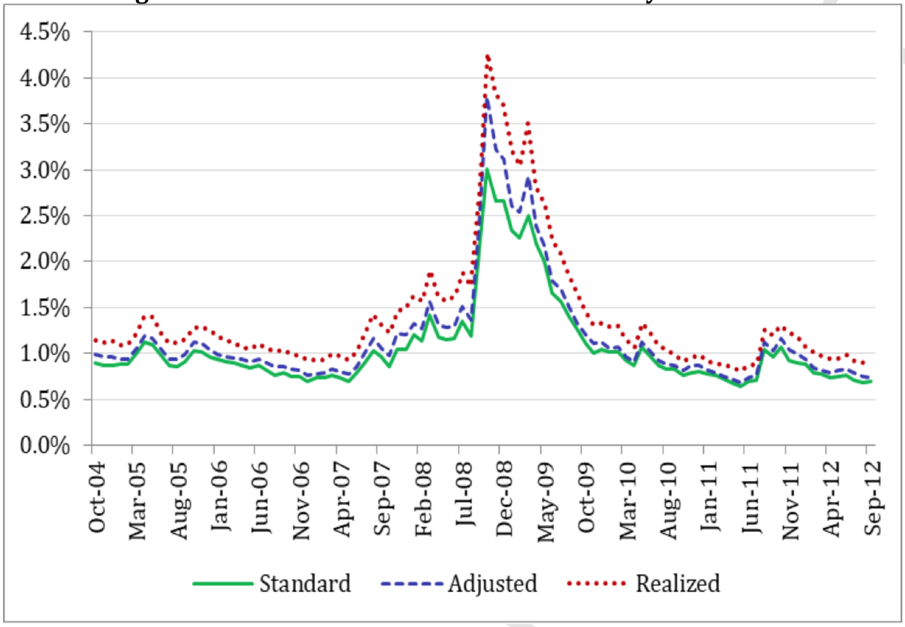
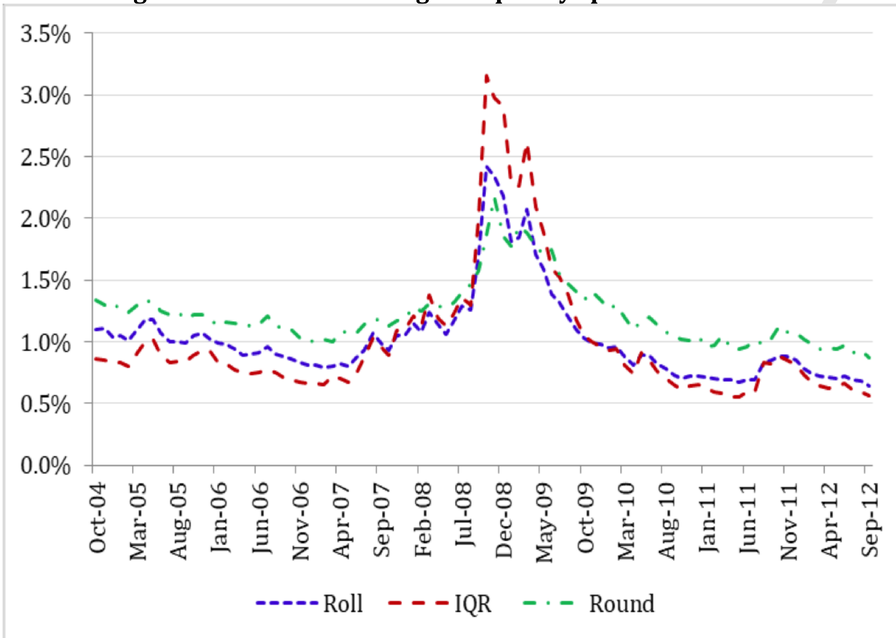


Figure 4. Time series of market-wide volatility estimators



This figure displays the monthly time series of the volatility estimators: the CS proposal (Standard), an adjusted version that accounts for infrequent trading (Adjusted), and the realized volatility (Realized), computed with daily returns in a window containing the last three months. The estimators are computed daily and for each bond during the period from October 1, 2004, to September 30, 2012. The monthly values are averaged over all days within the month and the figure shows the aggregate series that average across all bonds.

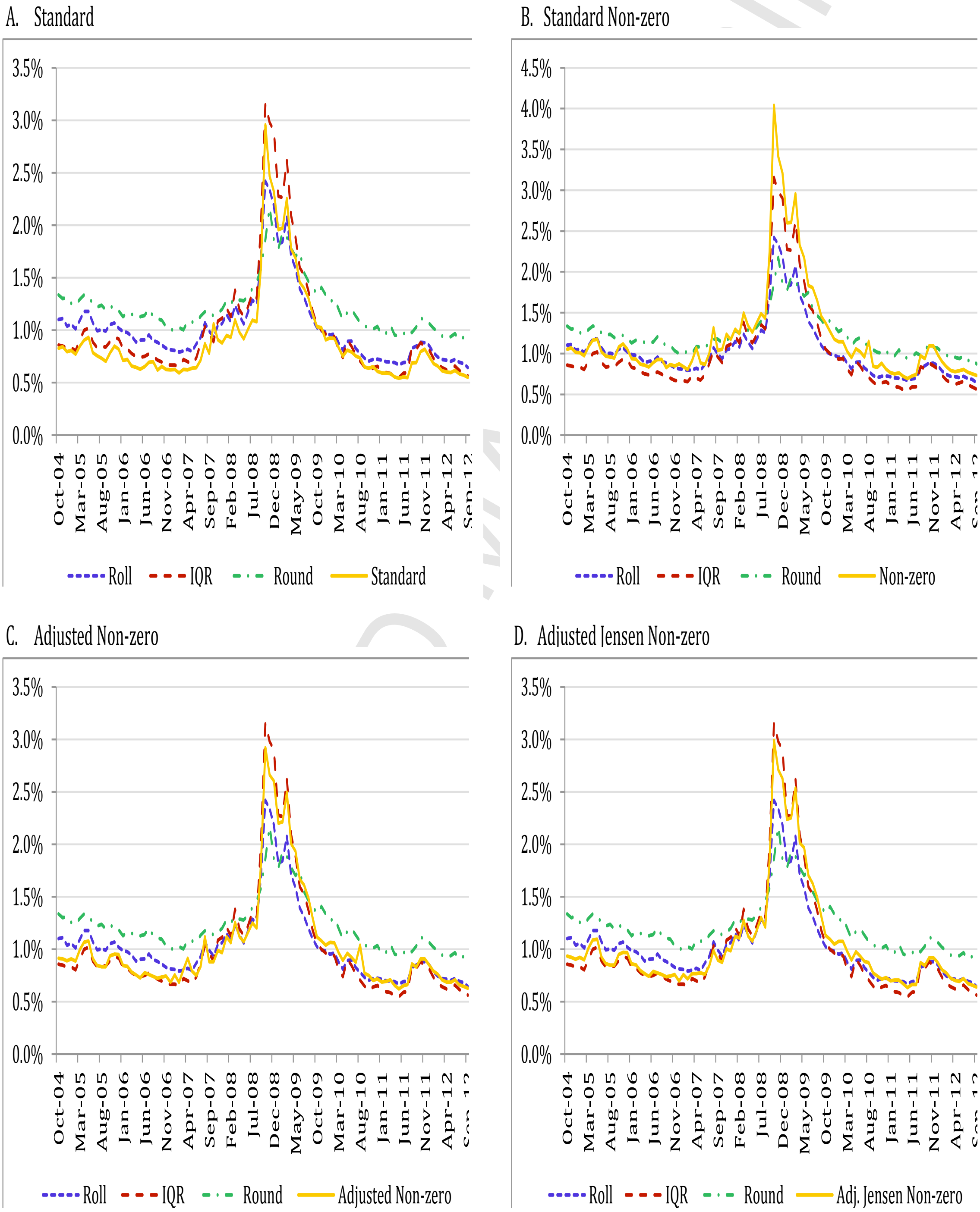
Figure 5. Time series of high-frequency spread estimators



This figure displays the monthly time series of the high-frequency transaction cost estimators computed from intraday prices. The measure Roll is the negative autocovariance in intraday returns, IQR is the interquartile range of trade prices within a day, and Round is the daily average of the round-trip cost computed with the high and low prices of each set of intraday round-trip trades. The estimators are computed daily and for each bond during the period from October 1, 2004, to September 30, 2012. The monthly values are averaged over all days within the month and the figure shows the aggregate series that average across all bonds.

Figure 6. Time series of high-frequency spread estimators and CS-based estimators

See the notes in Figure 5 for the high-frequency proxies' definitions and in Table3 for the CS-based estimators' definitions.



Footnotes

[1] Bao et al. (2011), Bongaerts et al. (2017), De Jong and Driessen (2012), Edwards et al. (2007), Felhütter (2012), and Friewald et al. (2012) are some examples.
[2] An additional problem is that the existence of the function is limited to $\sigma^2 \leq \beta/2(k_1 - k_2^2)$.
[3] Moreover, in the Internet Appendix of their paper, CS (2012) analyze the accuracy of their high–low estimator for intraday trade data with a 15-minute frequency and the three versions where intraday negative values are either included, set to zero, or excluded. The authors find that, although the three versions are downward biased in relation to the effective spread, the one that excludes negative cases produces the lowest absolute mean bias.
[4] After robustness checks, I find that this filter cleans better than alternative filters that use the mean instead of the median, a different number of standard deviations, or a different window length or the filter proposed by Rossi (2014).
[5] My sample is similar but not identical to that used by SSU, probably because the filters are in a different order and my outlier filter is slightly less restrictive.
[6] Note that SSU's bond characteristics and credit ratings come from Thomson Reuters and Bloomberg, while I work directly with FISD information. I use the average value between the four rating agencies, while SSU use three rating agencies and I provide the offering amount instead of the average outstanding amount over the life of the bond.
[7] The numerical computation of the volatility uses a grid of values between zero and the estimate that ignores Jensen's inequality plus 0.05, with changes of 0.00001, such that equation (4) or (16) holds for the standard and adjusted versions, respectively.
[8] In Table 3, the asterisks indicate rejection of the null that the means of the two compared measures are equal by applying the Wilcoxon rank sum test and the 1% significance level.
[9] Remember that cases with a negative spread are the same with and without Jensen's inequality.
[10] To compare the values for the standard liquidity proxy with those provided by SSU, I compute the cross-sectional statistics for the entire sample of bonds. The results for the mean, standard deviation, and three quartiles are 0.77, 0.63, 0.34, 0.61, and 1.04, respectively, while the numbers for SSU are 0.94, 0.95, 0.31, 0.63, and 1.25, respectively. However, SSU work with monthly series by averaging daily measures, while my statistics are computed directly from daily series. Additionally, the differences could be due to the samples of bonds not being entirely identical and from different sources (TRACE in my case and Bloomberg in theirs).
[11] The percentage of days with a negative spread estimate is the same for the standard and Jensen versions. In the case of the adjusted estimator, this percentage is a bit larger, only because the total number of days with an estimate is lower.
[12] Returns are computed only for days with trades and are divided by the number of days between the current and the previous observable price, such that all returns are on a daily basis.
[13] The exception is subsample 1, in which the highest correlations are obtained when *Roll* is used as the benchmark. Comparing between *Roll*, *IQR,* and *Round*, SSU also find the highest correlation with *IQR* in both the time series and cross-sectionally.